

PATENT ABSTRACTS OF JAPAN

(11) Publication number : 07-271521

(43) Date of publication of application : 20.10.1995

(51) Int.CI. G06F 3/06
 G06F 13/10

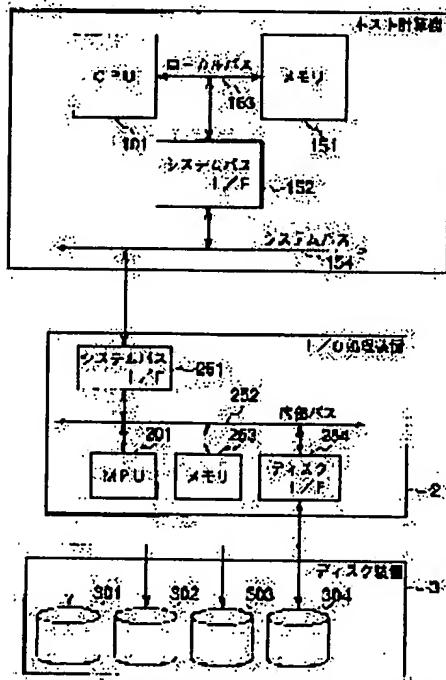
(21) Application number : 06-057197

(71) Applicant : HITACHI LTD

(22) Date of filing : 28.03.1994

(72) Inventor : MATSUNAMI NAOTO
 SUGA MASAYUKI
 KANEDA TAISUKE
 YAGISAWA IKUYA
 OEDA TAKASHI
 ARAKAWA TAKASHI

(54) COMPUTER SYSTEM



(57) Abstract:

PURPOSE: To reduce the interruption frequency to a CPU to report the end of the disk access.

CONSTITUTION: A CPU 101 of a host 1 issues plural disk commands to an I/O processor 2 by a disk command management means independently of and in relation to each other. An MPU 201 of the processor 2 registers the relation among received disk access commands and issues it to a disk device 3 by a disk command management part 203. Then the MPU 201 issues an interruption to the CPU 101 and informs a fact that all related commands are finished after this fact is decided by the end report received from the device 3.

LEGAL STATUS

[Date of request for examination] 01.07.1998

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3209634

[Date of registration] 13.07.2001

[Number of appeal against examiner's decision of
rejection]

[Date of requesting appeal against examiner's decision
of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(51) Int.Cl. ⁶ G 0 6 F 3/06 13/10	識別記号 5 4 0	序内整理番号 3 4 0 B 7368-5B	F I	技術表示箇所
--	---------------	---------------------------	-----	--------

審査請求 未請求 請求項の数13 O L (全 23 頁)

(21)出願番号 特願平6-57197	(71)出願人 株式会社日立製作所 東京都千代田区神田駿河台四丁目6番地
(22)出願日 平成6年(1994)3月28日	(72)発明者 松並 直人 神奈川県川崎市麻生区王禅寺1099 株式会社日立製作所システム開発研究所内
	(72)発明者 菅 政之 神奈川県海老名市下今泉810番地 株式会社日立製作所オフィスシステム事業部内
	(72)発明者 兼田 泰典 神奈川県川崎市麻生区王禅寺1099 株式会社日立製作所システム開発研究所内
	(74)代理人 弁理士 富田 和子

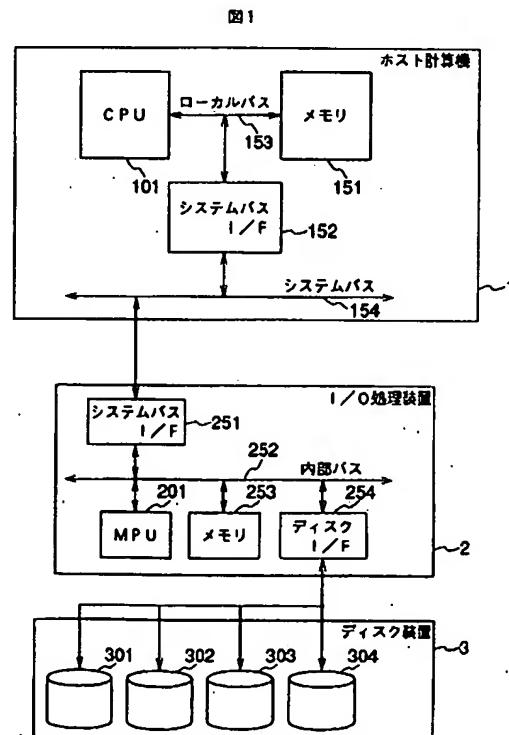
最終頁に続く

(54)【発明の名称】 計算機システム

(57)【要約】 (修正有)

【目的】ディスクアクセス終了を報告するCPUへの割込み回数を削減する。

【構成】ホスト1のCPU101は、ディスクコマンド管理手段で複数のディスクコマンドを関連付けて独立にI/O処理装置2に発行する。I/O処理装置2のMPU201は、ディスクコマンド管理部203で、受け取ったディスクアクセスコマンド間の関連を登録しディスク装置3に発行し、ディスク装置3からの終了報告より、関連付けられたすべてのコマンドが終了したことを判定すると、CPU101に割込みを発行し、関連付けられたコマンドすべてが終了したことを通知する。



【特許請求の範囲】

【請求項1】割込み処理機構を備えたCPUとメモリとを備えたホスト装置と、データの記録再生を行なう補助記憶装置と、ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置とを備えた計算機であって、

前記ホスト装置のCPUは、前記補助記憶装置のアクセスを指示する、複数のコマンドを、所定の条件に従いグループ化し、前記メモリに、各グループと当該グループに属するコマンドの対応を表す情報を記憶し、各コマンドに当該コマンドが属するグループを識別可能な識別情報を附加し、前記識別情報を附加した複数のコマンドを独立に前記入出力処理装置に渡す関連付け手段と、前記入出力処理装置より特定のグループについての処理が終了したことを割込みによって報告された場合に、前記メモリに記憶した各グループと当該グループに属するコマンドの対応を参照し、前記特定のグループに属する全てのコマンドに関する、所定の終了処理を行なう終了処理手段とを有し、

前記入出力処理装置は、渡された複数のコマンドに附加されている前記識別情報を、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶する手段と、前記CPUより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記CPUに割り込みを用いて報告する手段とを有することを特徴とする計算機システム。

【請求項2】割込み処理機構を備えたCPUとメモリとを備えたホスト装置と、補助記録媒体にデータの記録再生を行なう補助記憶装置と、補助記憶装置とホスト装置との間の入出力を担う入出力処理装置とを備えた計算機であって、

前記CPUは、前記補助記憶装置のアクセスを指示する、複数のコマンドを、所定の条件に従いグループ化し、前記メモリに、各グループと当該グループに属するコマンドの対応を表す情報を記憶し、前記複数のコマンドを独立に前記入出力処理装置に渡す関連付け手段と、前記入出力処理装置より特定のグループについての処理が終了したことを報告された場合に、前記メモリに記憶した各グループと当該グループに属するコマンドの対応を参照し、前記特定のグループに属する全てのコマンドに関する、所定の終了処理を行なう終了処理手段とを有し、

前記入出力処理装置は、前記CPUより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、前記メモリに記憶された各グループと当該グループに属するコマンドの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終

了した場合に、当該グループについての処理が終了したことを前記CPUに割込みを用いて報告する手段とを有することを特徴とする計算機システム。

【請求項3】請求項1または2記載の計算機システムであって、

前記補助記憶装置は、複数のディスク装置から構成されるディスクアレイ装置であって、

前記ホスト装置のCPUは、前記ディスクアレイ全体に対するアクセスを指示するディスクアレイコマンドにより、ディスクアレイ装置を構成する個々のディスク装置へのコマンドを生成する手段と、ディスクアレイコマンドと個々のディスク装置へのコマンドの対応を前記メモリに記憶する手段とを有し、

前記関連付け手段は、一つの前記ディスクアレイコマンドに対応する個々のディスク装置へのコマンドを、一つの前記グループとし、

前記終了処理手段は、前記特定のグループに属する全てのコマンドに関する、所定の終了処理として、前記特定のグループに属するコマンドに対応して前記メモリに記憶されているディスクアレイコマンドの終了処理を行なうことを特徴とする計算機システム。

【請求項4】請求項1または2記載の計算機システムであって、

前記補助記憶装置は、データを格納する複数のデータ用ディスク装置と、前記複数のデータ用ディスク装置の同ディスク内アドレスの領域に記録されているデータ全体のパリティを記憶する1以上のパリティ用ディスク装置から構成されるRAID型のディスクアレイ装置であつて、

前記ホスト装置のCPUは、前記ディスクアレイ全体に対する新データの書き込みアクセスを指示するディスクアレイコマンドより、新データの前記データ用ディスク装置への書き込みアクセスを指示する新データ書き込み用コマンドと、前記新データの書き込みアクセスにより更新される、前記データ用ディスク装置内の現データの読み出しアクセスを指示する現データ読み出し用コマンドと、前記新データの書き込みアクセスにより更新される、前記新データに対応する、前記パリティ用ディスク装置内の現パリティの読み出しアクセスを指示する現パリティ読み出し用コマンドと、前記新データの前記データ用ディスク装置への書き込みアクセスにより更新される、前記新データに対応する新パリティの前記パリティ用ディスク装置への書き込みアクセスを指示する新パリティ書き込み用コマンドを生成する手段と、

新パリティ生成するパリティ生成部とを有し、

前記関連づけ手段は、現データ読み出し用コマンドと前記現パリティ読み出し用コマンドとを一つの前記グループとし、現データ読み出し用コマンドと前記現パリティ読み出し用コマンドを前記入出力処理装置に渡し、その

後、前記パリティ生成手段によって、前記新パリティ書き込み用コマンドによって書き込む新パリティが生成されたら、新データ書き込み用コマンドと前記新パリティ書き込み用コマンドとを一つの前記グループとし、新データ書き込み用コマンドと前記新パリティ書き込み用コマンドとを前記入出力処理装置に渡し、

前記パリティ生成部は、前記入出力処理装置が、渡された現データ読み出し用コマンドに応じて前記ディスクアレイより読み出した前記現データと現パリティ用読み出し用コマンドに応じて前記ディスクアレイより読み出した現パリティと、新データ書き込み用コマンドによって書き込む新データとの排他的論理和を、前記新パリティ書き込み用コマンドによって書き込む新パリティとして生成することを特徴とする計算機システム。

【請求項5】請求項4記載の計算機システムであって、現データ読み出し用コマンドと前記現パリティ読み出し用コマンドとの前記グループと、新データ書き込み用コマンドと前記新パリティ書き込み用コマンドとの前記グループと、これらの二つのグループに対応するディスクアレイコマンドとの対応を示す情報を前記メモリに記憶する手段を有し、

前記終了処理手段は、前記特定のグループに属する全てのコマンドに関する、所定の終了処理として、前記メモリに前記特定のグループに対応して記憶されているディスクアレイコマンドに対応して記憶されている2つのグループについての処理が終了したことを、前記入出力処理装置より既に報告されているか否かを判定し、既に成されている場合に、当該ディスクアレイコマンドの終了処理を行なうことを特徴とする計算機システム。

【請求項6】請求項1または2記載の計算機システムであって、

前記ホスト装置のCPUは、前記補助記憶装置に書き込むデータを一旦、メモリ上に設けたキャッシュ領域に書き込む手段と、所定の契機で、前記キャッシュ領域に書き込まれた複数のデータの前記補助記憶装置への書き込みアクセスを指示する複数のコマンドを生成する書き戻し要求部とを、さらに有し、

前記関連付け手段は、前記書き戻し要求部が、前記所定の契機で生成した複数のコマンドの一部もしくは全部を、一つの前記グループとすることを特徴とする計算機システム。

【請求項7】請求項6記載の計算機システムであって、前記書き戻し要求部は、所定の条件が満たされた場合に、前記生成した、前記キャッシュ領域に書き込まれた複数のデータの前記補助記憶装置への書き込みアクセスを指示する複数のコマンドに、グループ化を禁止する情報を付加する手段を有し、

前記関連付け手段は、前記書き戻し要求部が、前記所定の契機で生成した複数のコマンドのうち、グループ化を禁止する情報が付されているコマンドについてはグル

ープ化の対象としないことを特徴とする計算機システム。

【請求項8】請求項1または2記載の計算機システムであって、

前記補助記憶装置は、データを格納する複数のデータ用ディスク装置と、前記複数のデータ用ディスク装置の同ディスク内アドレスの領域に記録されているデータ全体のパリティを記憶する1以上のパリティ用ディスク装置から構成されるRAID型ディスクアレイ装置であつて、

前記ホスト装置のCPUは、前記ディスクアレイ全体に対する新データの書き込みアクセスを指示するディスクアレイコマンドより、前記データ用ディスク装置への前記新データの書き込みアクセスを指示する新データ書き込み用コマンドと、前記新データの書き込みアクセスにより更新される、前記データ用ディスク装置内の現データの読み出しアクセスを指示する現データ読み出し用コマンドと、前記新データの書き込みアクセスにより更新される、前記データ用ディスク装置内のデータに対応する、前記パリティ用ディスク装置内の現パリティの読み出しアクセスを指示する現パリティ読み出し用コマンドと、前記新データの前記データ用ディスク装置への書き込みアクセスにより更新される、前記新データに対応する新パリティの前記パリティ用ディスク装置への書き込みアクセスを指示する新パリティ書き込み用コマンドを生成する手段と、

新パリティを生成するパリティ生成部とを有し、

前記関連付け手段は、現データ読み出し用コマンドを前記入出力処理装置に渡し、前記入出力処理装置によって現データが前記ディスクアレイより読み出されたら、現データ書き込み用コマンドを前記入出力処理装置に渡し、その後所定期間経過後に、複数の現パリティ読み出し用コマンドを、一つの前記グループとし、一つのグループとした複数の現パリティ読み出し用コマンドを前記入出力処理装置に渡し、前記パリティ生成手段によって書き込む新パリティが生成された、複数の新パリティ書き込み用コマンドを、一つの前記グループとし、一つのグループとした複数の新パリティ書き込み用コマンドを前記入出力処理装置に渡し、

前記パリティ生成部は、前記入出力処理装置が、渡された現データ読み出し用コマンドに応じて前記ディスクアレイより読み出した前記現データと、新データ書き込み用コマンドによって書き込む新データとの排他的論理和を算出する手段と、算出された現データと新データの排他的論理和を保持する手段と、現パリティ用読み出し用コマンドに応じて前記ディスクアレイより読み出した現パリティと保持したパリティの排他的論理和を、前記新パリティ書き込み用コマンドによって書き込む新パリティとして生成することを特徴とする計算機システム。

【請求項9】請求項8記載の計算機システムであって、前記ホスト装置のCPUは、前記ディスクアレイより読

み出したデータおよびパリティを、メモリ上に設けたキャッシュ領域に保持する手段を有し、前記関連付け手段は、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきデータに対応するデータが存在する場合には、現データ読み出し用コマンドを前記入出力処理装置に渡さず、前記キャッシュ領域に、現パリティ読み出し用コマンドによって読み出すべきパリティに対応するパリティが存在する場合には、現パリティ読み出し用コマンドを前記入出力処理装置に渡さず、

前記パリティ生成部は、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきデータに対応する出が存在する場合には、前記ディスクアレイより読み出した前記現データに代えて前記キャッシュ領域に存在するデータを用いて、前記現データと新データとの排他的論理和を生成し、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきパリティに対応するパリティが存在する場合には、前記ディスクアレイより読み出した前記現パリティに代えて前記キャッシュ領域に存在するパリティを用いて、現パリティと保持したパリティの排他的論理和を、前記新パリティ書き込み用コマンドによって書き込む新パリティとして生成することを特徴とする計算機システム。

【請求項10】ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置であって、ホスト装置より渡された複数のコマンドに付加されている、当該コマンドの属するグループを示す情報に基づいて、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶する手段と、ホスト装置より渡された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ホスト装置に割り込みを用いて報告する手段とを有することを特徴とする入出力装置装置。

【請求項11】ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置であって、前記入出力処理装置は、前記ホスト装置より渡されたコマンドの指示する補助記憶装置のアクセスを実行する手段と、前記ホスト装置に保持されたグループと当該グループに属するコマンドの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ホスト装置に割り込みを用いて報告する手段とを有することを特徴とする入出力処理装置。

【請求項12】割込み処理機構を備えたCPUを備えたホスト装置と、データの記録再生を行なう補助記憶装置と、ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置とを備えた計算機における、補助記憶装置

のアクセス方法であって、

前記ホスト装置のCPUにおいて、前記補助記憶装置のアクセスを指示する、複数のコマンドをグループ化するステップと、

前記ホスト装置のCPUにおいて、各コマンドに当該コマンドが属するグループを識別可能な識別情報を付加し、前記識別情報を付加した複数のコマンドを独立に前記入出力処理装置に渡すステップと、

前記入出力処理装置において、渡された複数のコマンドに付加されている前記識別情報を、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶するステップと、

前記入出力処理装置において、前記CPUより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行するステップと、

前記入出力処理装置において、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了したかを判定し、終了した場合に、当該グループについての処理が終了したことを前記CPUに割り込みを用いて報告するステップとを有することを特徴とする補助記憶装置のアクセス方法。

【請求項13】割込み処理機構を備えたCPUとメモリとを備えたホスト装置と、データの記録再生を行なう補助記憶装置と、ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置とを備えた計算機における、補助記憶装置のアクセス方法であって、

前記ホスト装置のCPUにおいて、前記補助記憶装置のアクセスを指示する、複数のコマンドを、所定の条件に従いグループ化し、前記メモリに、各グループと当該グループに属するコマンドの対応を表す情報を記憶するステップと、

前記複数のコマンドを独立に前記入出力処理装置に渡すステップと、

前記入出力処理装置において、前記CPUより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行するステップと、

前記メモリに記憶された各グループと当該グループに属するコマンドの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了したかを判定し、終了した場合に、当該グループについての処理が終了したことを前記CPUに割り込みを用いて報告するステップとを有することを特徴とする補助記憶装置のアクセス方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、主として、計算機システムにおける補助記憶装置へのアクセスの技術に関するものである。

【0002】

【従来の技術】従来、計算機システムにおいて補助記憶装置として用いられるディスク装置を高速化、高信頼化する技術として、複数台のディスク装置を1台のホスト計算機に接続するマルチディスクと呼ばれる技術が知られている。

【0003】たとえば、"A Case for Redundant Arrays of Inexpensive Disks(RAID)"InProc. ACM SIGMOD, June 1988(カルフォルニア大学 バークレー校発行)には、複数のディスク装置を専用の制御装置を介してCPUに接続し、複数のディスク装置へのデータの分配制御等のディスク管理をすべてこの制御装置にて実施することにより、CPUが、複数のディスク装置を1台のディスク装置に見なせるようにしている。

【0004】また、さらに、近年では、年率2倍弱に達するCPUの高性能化に伴い、専用の制御装置を用いずに、複数のディスク装置の制御をCPUによりソフトウェア的に実施する技術、いわゆるソフトアレイ、またはソフトRAIDと呼ばれる技術が実現されている。

【0005】この技術によれば、CPUのディスクインターフェースに複数台のディスク装置を接続する。そして、CPU上で実行されるドライバプロセスが、他のプロセスが、複数台のディスクを仮想的に一台のディスク装置に見なして発行したコマンドを、個々のディスク装置へのコマンドに変換して個々のディスク装置に発行する。また、ドライバプロセスは、記録時にはデータの誤り訂正用の冗長データの生成し、データと共にディスク装置に記録する。また、再生時には、再生データの誤りを、共に記録した誤り訂正用の冗長データを用いて訂正する。

【0006】このように、ソフトアレイ、またはソフトRAIDと呼ばれる技術によれば、年々高速化が進むCPUの能力を有効に利用して、低価格で、かつ高性能、高信頼なディスクシステムを実現することができる。

【0007】ところで、ディスク装置のディスクへのアクセスは、機械的な動作が必要であることから時間がかかる。

【0008】そこで、従来より、ディスク装置のディスクへのアクセスを、見かけ上、早めるために、従来、リードキャッシング、ライトキャッシングと呼ばれる技術が広く用いられている。

【0009】これらの技術では、主記憶上にキャッシング用の領域を確保する。そして、リードキャッシングによれば、ディスクからデータをリードした際にキャッシング用の領域に、そのデータを保管し、もし、つぎに同じデータのリード要求がプロセスより発行された際には、ディスクアクセスを行なうこと無く主記憶のキャッシング用領域からデータをリードし要求元のプロセスに与えることにより、ディスクアクセス回数を削減する。また、ライトキャッシングによれば、ライト時に主記憶上のキャッシング用の領域にデータを書き込み、ディスクアクセス要求

元のプロセスに終了通知を発行する。そして、ある時間後、主記憶上のキャッシング用の領域のデータをまとめてディスクにライトすることでディスクアクセス回数を削減する。

【0010】これらのリードキャッシング、ライトキャッシングの技術は、ディスクアクセス要求元のプロセスの、アクセスの時間的、空間的局所性を利用したもので、効果が大きいため、多くの計算機システムにおいて広く用いられている。

【0011】

【発明が解決しようとする課題】さて、計算機システムにおいて、ユーザのプログラムは、一般にOS(オペレーティングシステム)という基本プログラムの管理のもとで動作している。OSは、プログラムを並列に実行させる。この並列動作を実現するため、OSはプログラムをプロセスという単位で管理し、これを時分割に切り替えながら動作させている。

【0012】また、一般的な計算機システムでは、CPUから、ディスク装置にディスクをアクセスするコマンドが発行され、ディスク装置がアクセスを終了すると、ディスク装置から前記コマンドに対応するアクセスの終了を通知するための割込みがCPUに送られる。CPUは、この割込みを受け取ると、その時動作しているプロセスを一旦停止し、ディスクアクセス終了のための所定の割込み処理を開始する。

【0013】ここで、この割込み処理は、次のような手順により行なわれる。すなわち、CPUは、まず、動作中のプロセスを一旦停止し、このプロセスを再度動作させるために必要な全ての環境を保存し、割り込み処理を実行するのに必要な環境を準備し、所定の割り込み処理を実施し、割り込み処理を実施する以前の環境に戻し、停止したプロセスを再起動する。

【0014】このように、割込み処理の実行は、CPUにとって負荷が大きい。

【0015】一方、前述したソフトアレイ、またはソフトRAIDと呼ばれる技術によれば、CPU上で実行されるドライバプロセスが、個々のディスク装置への複数のコマンドに変換して個々のディスク装置に発行する。したがい、ディスク装置よりの、アクセスの終了を通知するための割込みも、個々のディスクに発行した複数のコマンドのそれぞれについて発生する。

【0016】また、高信頼化のために誤り訂正用の冗長データを生成し、データと共にディスクに保存する場合には、CPUは、この冗長データを生成するために必要なデータのリードコマンドや、冗長データのライトコマンドを、ユーザプロセスより要求されたデータのアクセスコマンドに加えて発行するため、さらに、ディスク装置よりの、アクセスの終了を通知するための割込みの発生回数が増加する。

【0017】すなわち、前述したソフトアレイ、または

ソフトRAIDと呼ばれる技術によれば、割込み処理の発生頻度が高く、CPUの、込み処理の実行のための負荷が過大となってしまい、ユーザプログラムを十分なスピードで実行できなくなったり、マルチディスクの能力を充分に引き出せなくなってしまう。

【0018】一方、前記ライトキャッシュの技術によれば、ある時間毎に、主記憶上のキャッシュ用領域から、まだディスクに書き戻されていないデータを書き出す処理を行う必要がある。また、主記憶の容量には限りがあるため、ディスクに書き戻していないデータ量が一定量を超えたら、次のデータの書き込みにキャッシュ用領域を用いることができるよう、ある一定以下の容量に減少するまでデータのディスクへの書き出す処理を行なう必要がある。

【0019】また、これらの書き出し処理では、どちらも一度にたくさんデータをディスクに書き込む必要があるので、CPUは、このような書き出し処理のために多くのライトコマンドを、ユーザプロセスより要求されたデータのアクセスコマンドに加えて発行する。このため、ディスク装置よりの、アクセスの終了を通知するための割込みの発生回数が増加する。

【0020】すなわち、ライトキャッシュの技術でも、割込み処理の発生頻度が高くなるために、CPUの、込み処理の実行のための負荷が過大となってしまう。

【0021】そこで、本発明は、ディスクアクセスに起因する、CPUへの割り込み発生回数を削減することにより、CPUの負荷を低減することのできる計算機システムを提供することを目的とする。

【0022】

【課題を解決するための手段】前記目的達成のために本発明は、割込み処理機構を備えたCPUとメモリとを備えたホスト装置と、備えた補助記録媒体にデータの記録再生を行なう補助記憶装置と、ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置とを備えた計算機であって、前記ホスト装置のCPUは、前記補助記憶装置のアクセスを指示する、複数のコマンドを、所定の条件に従いグループ化し、前記メモリに、各グループと当該グループに属するコマンドの対応を表す情報を記憶し、各コマンドに当該コマンドが属するグループを識別可能な識別情報を付加し、前記識別情報を付加した複数のコマンドを独立に前記入出力処理装置に渡す手段と、前記入出力処理装置より特定のグループについての処理が終了したことを割込みによって報告された場合に、前記メモリに記憶した各グループと当該グループに属するコマンドの対応を参照し、前記特定のグループに属する全てのコマンドに関する、所定の終了処理を行なう終了処理手段とを有し、前記入出力処理装置は、渡された複数のコマンドに付加されている前記識別情報より、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶する手段と、前記CPUより

独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記CPUに割り込みを用いて報告する手段とを有することを特徴とする計算機システムを提供する。

【0023】なお、前記入出力処理装置に代えて、前記CPUより独立に発行された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、前記メモリに記憶された各グループと当該グループに属するコマンドの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記CPUに割り込みを用いて報告する手段とを有する入出力処理装置を用いるようにしてもよい。

【0024】

【作用】本発明に係る計算機システムによれば、個々のコマンドについての処理が終了する毎に前記CPUに報告する代わりに、各コマンドと当該コマンドが属するグループの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了した場合にのみ、当該グループについての処理が終了したことを前記CPUに割り込みを用いて報告する。したがい、CPU割り込み処理の回数を削減でき、CPUの負荷を軽減できる。

【0025】

【実施例】以下、本発明に係る計算機システムの実施例について説明する。

【0026】図1に、本実施例に係る計算機システムの構成を示す。

【0027】図中、1はホスト計算機（以下、「ホスト」と呼ぶ）、101はホスト1の全体制御、各種計算を実行するCPU、151は、CPU101が実行するプログラムや、データ等を格納するメモリ、153はCPUのバスであるローカルバス、154は周辺装置等を接続するためのシステムバス、152はローカルバス153とシステムバス154を接続し、CPU101から周辺回路を接続するためのシステムバスインターフェース（以下、「システムバスI/F」と記す）である。

【0028】また、2はホスト1のシステムバス154に接続するディスク装置の入出力制御を司るI/O処理装置、201はI/O処理装置2の全体制御、および各種演算を実行するMPU、251はホスト1のシステムバス154に接続するためのシステムバスインターフェース（以下、「システムバスI/F」と記す）、252はI/O処理装置の内部バス、253はMPUが実行するプログラムや、必要なデータ等を格納するメモリ、254はディスクを接続するためのディスクインターフェース（以下、「ディスクI/F」と記す）である。

【0029】3はディスク装置であり、複数台のハードディスク装置（以下、「ディスク」という）（301～305）から構成されるディスクアレイである場合もあり、1台もしくは複数台の単体のディスクである場合もある。本第1第実施例では、1台もしくは複数台の単体のディスクで構成されるものとして説明する。

【0030】ここで、まず、CPU101のディスクアクセス動作におけるデータの流れを説明する。

【0031】ホスト1において、CPU101は、プログラム実行中に、ディスク3からのデータリードが必要になったとすると、システムバスI/F152を介してI/O処理装置2にリードコマンドを発行する。I/O処理装置2では、システムバスI/F251を介して、MPU201が上記リードコマンドを受信し、これを解析し、対象となるディスク3へのコマンド（ディスクコマンド）へ変換し、ディスクI/F254を介し、対象ディスク3にコマンドを発行する。ディスク3はこのディスクコマンドを受信し、ヘッドの位置決め後、対象データをI/O処理装置2に転送する。I/O処理装置2はディスクI/F254を介してこのデータを受信し、システムバスI/F251、およびホストのシステムバス154、システムバスI/F152経由でホスト1のメモリ151に転送する。このデータ転送はホスト1のCPU101が直接I/O処理装置のシステムバスI/F251からデータをリードし、メモリ151に転送する方式と、I/O処理装置2が独自にホスト1のメモリ151に転送する方式とがあるが、本実施例では、I/O処理装置2が独自にホスト1のメモリ151に転送するものとする。

【0032】以上の処理により、ディスク3からのデータのリード処理が完了し、ホスト1のCPU101は、このデータを用い演算を実施できる。なお、ここではリード動作について説明したが、ライト時にはデータ転送方向が逆になるのみで同様である。

【0033】次に、このような、CPU101のディスクアクセス動作を実現する、ホスト1のCPU101と、I/O処理装置2のMPU201の構成を、コマンド処理系を中心に図2に示す。

【0034】図中、1010はCPU101上で実行されるオペレーティングシステム(OS)、102はディスク上に格納しているファイルのアクセス要求を生成、発行するユーザプロセス、1020はキャッシュデータライト要求部、1000はディスクドライバ部である。

【0035】なお、ユーザプロセス102、キャッシュデータライト要求部1020、ディスクドライバ部1000は、実際には、CPU101で実行されるプログラムであるオペレーティングシステム1010の制御下で、CPU101で実行されるプロセスとして実現される。

【0036】なお、図では、CPU101とMPU201

1を直接接続して示しているが、実際には、CPU101とMPU201間の接続は、システムバスI/F152、251を介して行なわれる。また、MPU201とディスク装置3を直接接続して示したが、実際には、MPU201とディスク装置3間の接続は、ディスクI/F254を介して行なわれる。

【0037】ディスクドライバ部1000中の、103はファイルとディスク上のデータ記録アドレスとの対応づけ管理を行い、ユーザプロセス102、キャッシュデータライト要求部1020により生成、発行されたファイルへのアクセス要求を、実際のディスクへのアクセス要求に変換するファイルアクセス要求管理部、104は前記ファイルアクセス要求管理部103により発行された、ディスクへのアクセス要求を受信し、各ディスク3へのコマンドを生成し、後述する関連するコマンドの関連付け処理や、コマンドの発行処理、終了処理を行なうディスクコマンド管理部、105は前記ディスクコマンド管理部104により生成されたディスクコマンドをI/O処理装置2に発行し、またディスクコマンド終了通知を受信するI/O処理装置制御部、106は、メモリ151上に設けられたキャッシュ用領域であるディスクキャッシュ部である。

【0038】また、前記ディスクコマンド管理部104中の、1040は前記ファイルアクセス要求管理部103から発行されたディスクアクセス要求を受信するディスクアクセス要求受信手段、1041は終了したディスクアクセス要求について前記ファイルアクセス要求管理部に終了報告を発行するディスクアクセス要求終了報告発行手段、1042は前記ディスク装置3がディスクアレイである場合に、発行されたアクセス要求を複数台の各ディスクへのアクセス要求に変換するディスクアレイ管理手段、1043はディスク装置の管理と、後述するディスクコマンドに関連付けを施すディスクコマンド関連付け管理手段10430を内蔵するディスクコマンド管理手段である。

【0039】また、1044は、前記ディスクコマンド関連付け管理手段1043により関連付けられたコマンド群の関連情報を管理する関連コマンド管理テーブル等（後に詳述する）を記憶する、メモリ151上に設けたコマンド管理エリアである。

【0040】次に、MPU201中、2010はOSに相当する全体制御部、202は前記ホスト1のI/O処理装置制御部105が発行したディスクコマンドを受信し、全体制御部2010に通知し、ディスクコマンドの終了をホスト1に通知するI/O処理装置I/F部、203は前記受信したコマンドの登録・終了管理、および、関連付けられたコマンドの登録・終了管理を行なうディスクコマンド管理部、204は前記ディスクコマンドに従いディスクにアクセスしディスク装置とホスト1のメモリ151間のデータ転送を実行するディスクコマン

ド処理部である。なお、I/O処理装置1/F部202、ディスクコマンド管理部203、ディスクコマンド処理部204は、実際には、MPU201上で実行されるプログラムである全体制御部2010の制御下で実行されるプロセスである。

【0041】また、205は、メモリ253上に設けたコマンド管理エリアであり、後述する関連コマンド管理テーブル等を格納する。

【0042】次に、ディスクコマンド管理部203中の、2030は前記ホスト1内部のディスクコマンド管理手段1043により関連付けられたコマンドの関連付けの内容を、I/O処理装置2内のメモリ253のコマンド管理エリア205に登録する関連コマンド登録手段、2031は登録した関連コマンドに対応するディスクアクセスがすべて終了したかどうかを判定する関連コマンド終了判定手段である。

【0043】以下、本第1実施例におけるディスクアクセス動作について説明する。

【0044】図2において、実行されているユーザプロセス102においてディスク上のデータのリードライトが必要になると、ファイルアクセス要求をOS1010に発行する。OS1010は、ディスクドライバ部1000を起動し、ファイルアクセス要求を渡す。

【0045】ディスクドライバ部1000の、ファイルアクセス要求管理部103は、内部に備えているファイル格納情報テーブルを参照して、ファイルアクセス要求に従い、アクセスすべきディスク3と、ディスク3上のファイルの転送先頭アドレスと、ディスク3へのデータ転送長を得、これらを含むディスクアクセス要求にファイルアクセス要求を変換し、ディスクコマンド管理部104に送出する。

【0046】ディスクコマンド管理部104のディスクアクセス要求受信手段1040は、このディスクアクセス要求を受信し、ディスクコマンドパケットをメモリ151上に生成する。

【0047】ディスクコマンドパケットの構成は図3に示すとおりであり、コマンドパケットCPには、(1)使用するI/O処理装置(ボード)へのコマンドの種類、(2)対象(ターゲット)ディスクの番号、(3)対象ディスク中の論理装置番号(パーティション番号、論理ユニット番号等)、(4)データ転送長、(5)対象データのディスク上のアドレス、(6)ディスクコマンドの種類(リード/ライト)、(7)データの転送先(元)メモリアドレス、(8)コマンドパケットの実行ステータス情報、(9)ホスト内部またはI/O処理装置内部で複数のコマンドパケットをキューイングする際にキュー管理に用いるコマンドキューイングポインタ(以上(1)から(9)を内容Aと称する)、(10)関連付けコマンドの管理に必要なコマンド付加パケット管理用のポインタ、等のフィールドを有している。

【0048】ディスクアクセス要求受信手段1040は、このような1つのコマンドパケットの領域をメモリ151上に確保し、(1)使用するI/O処理装置(ボード)へのコマンドの種類、(2)対象(ターゲット)ディスクの番号、(3)対象ディスク中の論理装置番号(パーティション番号、論理ユニット番号等)、(4)データ転送長、(5)対象データのディスク上のアドレス、(6)ディスクコマンドの種類(リード/ライト)、(7)データの転送先(元)メモリアドレス等のフィールドに情報を格納した後、ディスクアレイ管理手段1042、にそのコマンドパケットのメモリ151上のアドレスを渡す。

【0049】ここで、本第1実施例では、ディスク装置3は単体のディスクより構成されているものとしているので、ディスクアレイ管理手段1042は、何もせずに、受け取ったコマンドパケットのアドレスをディスクコマンド管理手段1043に渡す。ディスク装置3がディスクアレイ装置である場合のディスクコマンド管理手段1043の動作については、後に第2実施例として説明する。

【0050】ここで、次にコマンドパケットのアドレスを受け取るディスクコマンド管理手段1043の構成を図4に示す。

【0051】図中、10431は、ホスト1に接続されているI/O処理装置2がディスクコマンドの関連付け機能をサポートしているかどうかの判断をするディスクコマンド関連付け機能サポート有無判定手段、10432はディスクコマンドパケットが既に関連付けがなされているかどうかを判定するディスクコマンド関連既所持有無判定手段、10433はコマンドパケットアドレスをキューイングするためのキューイング手段、10434はディスク資源の管理を行うディスク装置管理手段、10430はディスクコマンドに関連付けを施し、また、終了した関連ディスクコマンドの終了処理を行うディスクコマンド関連付け管理手段である。

【0052】このような構成において、ディスクコマンド管理手段1043はコマンドパケットアドレスを受信すると、ディスクコマンド関連付け機能サポート有無判定手段10431は、このメモリ151の、このアドレスのコマンドパケットを参照し、コマンドパケットの対象とするディスク装置の接続しているI/O制御装置が、後に説明するような関連付け機能をサポートしているかどうかを判断する。なお、この判断のために、あらかじめ、ホスト1に接続している各I/O処理装置2についての情報をディスクコマンド管理部104に登録しておくようとする。

【0053】さて、この判断により、関連付け機能をサポートしていないディスク装置へのコマンドパケットであることが判明したならば、このコマンドパケットのアドレスは直接ディスク装置管理手段10434へ送出される。もし、サポートしているならば、コマンドパケッ

トのアドレスは、次にディスクコマンド関連既所持有無判定手段10432に送出される。

【0054】ディスクコマンド関連既所持有無判定手段10432は、受け取ったアドレスのディスクコマンドパケットが既に関連付けがなされているかどうかを判定する。ディスクコマンドの関連付けは、後に第2実施例で説明するように、上位に位置するディスクアレイ管理手段1040によって関連付けられている場合があるからである。ディスクコマンド関連既所持有無判定手段10432は、既に関連付けが成されている場合は、再度関連付けを施す必要はないため、ディスク装置管理手段10434に、コマンドパケットのアドレスを送出する。

【0055】一方、ディスクコマンド関連既所持有無判定手段10432は、コマンドパケットに関連付けがまだなされていない場合には、コマンドパケットのアドレスをキューイング手段10433に送出する。キューイング手段10433は、到着したコマンドパケットを、順次、メモリ151上に設けたキューに格納する。このとき、コマンドパケットのアドレスが格納されたという情報はディスクコマンド関連付け管理手段10430に通知される。

【0056】以上で、一つのファイルアクセス要求に対するディスクドライバ部1000の処理は終了し、CPU101上では、他のプロセスが実行される。

【0057】一方、ディスクドライバ部1000は、OS1010の管理するタイマのタイマ割込みや、OS1010上で動作する所定のディスクコマンド管理プロセス102'からの起動命令によって、適当な時間間隔で定期的に起動される。そして、タイマ割込みによって起動された場合には、以下に説明する処理を実行する。なお、このように、定期的にディスクドライバ部1000を起動するのではなく、キューイング手段10433より通知された情報より、キューに格納されたコマンドパケットのアドレス数が一定数になった場合に、定期的に起動された場合に行なうものとして説明する以下の処理を、引き続き実行するようにしてもよい。

【0058】さて、タイマ割込みによって起動された場合、ディスクドライバ部1000のディスクコマンド関連付け管理手段10430は、キューに格納されているいくつかのコマンドパケットのアドレスの示すメモリ151上のコマンドパケットを走査し、関連付け可能なコマンドパケットの群を選びだす。関連付け可能なコマンドパケットの判定基準は、任意に設けることができる。なお、本明細書では、後述する第2実施例以降の実施例でいくつかの具体例を提示する。

【0059】そして、もし、関連付け可能ないいくつかのコマンドパケットを検出したならば、ディスクコマンド関連付け管理手段10430はキューイング手段10433から、順次、これらのコマンドパケットのアドレス

を取り出す。そして、これらのコマンドパケットに関連付けを施す。

【0060】ここで、この関連付けの詳細について説明する。

【0061】図3は、3つのコマンドパケットCP1、CP2、CP3が関連付けの対象として選ばれた場合を示している。いま、この選ばれたコマンドパケットを関連コマンドパケットと呼ぶことにする。また、選ばれたコマンドパケットの集合を関連コマンドパケット群と呼ぶことにする。

【0062】ディスクコマンド関連付け管理手段10430は、まず、これら各々の関連コマンドパケットに対応するコマンド付加パケットをメモリ151上に作成し、コマンドパケット中の関連付けコマンド付加パケットポインタに、コマンド付加パケットのアドレスを格納することによりコマンド付加パケットをコマンドパケットにリンクさせる。

【0063】ここで、コマンド付加パケットは、(1)関連コマンド群に固有の識別番号(関連コマンドタグ)、(2)関連コマンド群中のコマンドパケットの数、(3)関連コマンドパケットの実行ステータス情報(以上、(1)から(3)を内容Bと称する)、(4)対応する関連コマンド管理テーブルへのポインタ、等のフィールドを有している。

【0064】また、ディスクコマンド関連付け管理手段10430は、一つの関連コマンドパケット群に対して、1つの関連コマンドテーブル1044をメモリ151上に作成する。関連コマンド管理テーブルは、(1)関連コマンド群の固有の識別番号(関連コマンドタグ)、(2)関連コマンドパケット群中のコマンドパケットの総数、(3)関連コマンドパケット群中の未終了のコマンドパケット数、(4)関連コマンド群の実行ステータス(以上(1)から(4)を内容Cと称する)、(5)次に述べるコマンドパケットアドレスポインタテーブルを管理するためのリンクポインタ、等のフィールドを有している。

【0065】また、ディスクコマンド関連付け管理手段10430は、コマンドパケット毎に、コマンドパケットアドレスポインタテーブルを作成する。コマンドパケットアドレスポインタテーブルには、(1)コマンドパケットへのアドレスポインタ、(2)次のコマンドパケットアドレスポインタテーブルへのポインタ、等のフィールドを設ける。

【0066】このように、コマンドパケットおよび関連コマンド管理テーブルは、コマンド付加パケット、マンドパケットアドレスポインタテーブルを介して双方向で参照可能な構成となっている。

【0067】さて、以上のように、ディスクコマンド関連付け管理手段10430は関連付け可能なコマンドパケットを抽出し、メモリ上のコマンド管理エリア1044上の関連コマンド管理テーブルに登録する。

【0068】次にディスクコマンド関連付け管理手段10430はディスク装置管理手段10434にコマンドパケットのアドレスを発行する。

【0069】ディスク装置管理手段10434は、渡されたアドレスのコマンドパケットに、コマンド付加パケットがリンクされている場合は、このコマンド付加パケットを付加して、I/O処理装置制御部105に発行する。

【0070】次に、I/O処理装置制御部105は、このコマンドパケットを受信し、I/O処理装置2に、コマンドパケットを送出する。具体的には、I/O処理装置2のシステムI/F部251のコマンドレジスタにコマンドパケットを書き込む。

【0071】以上の処理が終了したらディスクドライバ部1000は処理を処理し、CPU101上では、他のプロセスが実行される。

【0072】一方、I/O処理装置2では、システムI/F部251のコマンドレジスタに書き込まれたコマンドパケットは、I/O処理装置I/F部202のディスクコマンド受信手段2020により、その後メモリ253に書き込まれ、その書き込みアドレスが、ディスクコマンド管理部203に送出される。

【0073】ディスクコマンド管理部203はこのアドレスを受信し、関連コマンド登録手段2030で、このアドレスのコマンドパケットが関連コマンドの1つのコマンドパケットであるか否かを付加されているコマンド付加パケットより判定する。

【0074】そして、関連付けがなされているならば、メモリ253上のコマンド管理エリア205の関連コマンドテーブルを参照し、コマンド付加パケットと同じ関連コマンドタグを持った関連コマンド管理テーブルの有無を確認し、もし、存在するならば、コマンド付加パケットの関連コマンド管理テーブルアドレスに、この関連コマンド管理テーブルのアドレスを登録し、さらに、コマンドパケットに対応するコマンドパケットアドレスポイントテーブルを作成し、関連コマンド管理テーブルにリンクする。また、もし、コマンド付加パケットと同じ関連コマンドタグを持った関連コマンド管理テーブルが存在しないならば、新規に関連コマンド管理テーブルを生成し、コマンド付加パケットの関連コマンド管理テーブルアドレスに、この関連コマンド管理テーブルのアドレスを登録し、さらに、コマンドパケットに対応するコマンドパケットアドレスポイントテーブルを作成し、関連コマンド管理テーブルにリンクする。ここで、新規登録時には、関連コマンド管理テーブルの関連コマンドパケット数並びに残コマンド数はコマンド付加パケット中の関連コマンドパケット数を設定しておく。

【0075】このようにして、ディスクコマンド管理部203は、メモリ151に作成された図3の情報と、同等の情報をメモリ253上に構築する。

【0076】さて、この後、ディスクコマンド管理部203は、メモリ253上に設けられている対象ディスクの管理テーブルより、コマンドパケットの対象ディスクの稼動状況を判断し、もし対象ディスクが稼動中であればメモリ253上に設けたコマンドキューに、このコマンドパケットのアドレス登録し、ディスクが空き状態になるのを待つ。もし、対象ディスクが空き状態であるならば、対象ディスクの管理テーブルに稼動中であることを設定し、この管理テーブルにコマンドパケットのアドレスを登録する。また、ディスクコマンド処理部204に、コマンドパケットのアドレスを渡す。

【0077】ディスクコマンド処理部204は、渡されたアドレスのコマンドパケットに従い、ディスクにアクセスし、ホスト1のメモリ151とディスク間のデータ転送を行ない、これが終了したら、終了報告をディスクコマンド管理部203に送る。

【0078】ディスクコマンド管理部203の関連コマンド終了判定手段2031は、前記ディスク管理テーブルを参照して、終了報告に対応するコマンドパケットを見つけだし、関連コマンドパケットであるかどうかを判定し、もし、関連コマンドパケットあるならば、コマンドパケットに付加されているコマンド付加パケットのフィールド「関連コマンド管理テーブルアドレス」から関連コマンド管理テーブルを参照し、そのフィールド「残コマンド数」の値を1減じた値とする。そして、この引算処理後、残コマンド数が0でないならば(>0)、同じ関連コマンドパケット群に属する全ての関連コマンドパケットの処理は終了してはいないので、コマンド付加パケットのフィールド「実行ステータス」の値を『コマンド終了待機中』に設定し、何もせずこのまま待機する。もし、引算処理後、残コマンド数が0であるならば、同じ関連コマンドパケット群に属する全ての関連コマンドパケットの処理は終了しているので、ディスクコマンド管理部203はI/O処理装置I/F部202に、終了した関連コマンドパケット群のフィールド「関連コマンドタグ」と、「実行ステータス」を関連コマンド管理テーブルより読み出し通知する。通知を受けた、I/O処理装置I/F部202のディスクコマンド終了報告発行手段2021は、ホスト1に、関連コマンドパケット群の処理終了を報告する割込み信号206をCPU101に送出する。

【0079】割り込み信号206はCPU101のハードウェアである割込み処理機構を介してOS1010に伝えられる。OS1010は、その時、処理中のプロセスを一旦停止し、CPU1内部のレジスタ内容等の状態情報(コンテキスト)を保存し、ディスクドライバ部1001を起動する。

【0080】割り込み信号206に基づき起動された、ディスクドライバ部1001の、ディスクコマンド終了受信手段1051は、I/O処理装置2のI/O処理装

置 I/F 部 202 から、関連コマンドタグと、関連コマンドパケット群の実行ステータスを獲得し、関連コマンドパケット群が正常終了したことを確認した後、ディスクコマンド管理手段 1043 にこの獲得した関連コマンドタグと実行ステータスを送出する。

【0081】CPU101 のディスクドライバ部 1000 のディスクコマンド管理手段 1043 (図 4 参照) は、終了した関連コマンドタグを受信し、ディスクコマンド関連付け管理手段 10430 で、関連コマンドタグにより関連コマンド管理テーブル 1044 を参照し、終了した関連コマンドパケットの適切な終了処理を実施し、関連コマンド群に属する個々のコマンドの終了報告を、ディスクアレイ管理手段 1042、ディスクアクセス要求終了報告発行手段 1041、ファイルアクセス要求管理部 103 を介し、ファイルアクセス要求の発行元に終了報告を発行する。また、同時に、関連コマンドタグに対応する関連コマンド管理テーブル 1044 を削除し、当該関連コマンドパケット群に関するすべての処理を終了する。

【0082】一方、関連コマンド群の実行ステータスを CPU101 側に渡した I/O 処理装置 2 の I/O 処理装置 I/F 部 202 は、ディスクコマンド管理部 203 に関連コマンドタグを戻し、ディスクコマンド管理部 203 は戻された関連コマンドタグを有する関連コマンド管理テーブル 205 と、これにリンクする関連コマンドパケットコマンドアドレスポインタテーブルと、戻された関連コマンドタグを有する関連コマンドパケットを付けられているコマンド付加パケットと共に削除し、この関連コマンド群に関する全処理を終了する。

【0083】以上、本第 1 実施例に係る計算機システムの動作を説明した。このような動作によって、実現されるコマンドと、これに対する報告のシーケンスを図 5 に示しておく。図 5 の、横方向は時間の経過を表し、縦軸方向の矢印は、各部位間で渡されるコマンド、報告を表している。

【0084】図示するように、複数のディスクコマンドを 1 つの関連コマンド群として管理し、CPU101 と I/O 処理装置 2 での協調動作することで、従来の n 個のディスクコマンドに対応する n 個の I/O 処理装置からホストへの終了報告を、唯一回の終了通知に削減することができる。

【0085】さて、前述したように、一般に、この終了報告を行う際には、I/O 処理装置からホストへ割り込み信号を発行し、CPU がこれを受信し、現在動作中のプログラム (プロセス) を停止し、割り込み処理を実行するが、この割込み処理に CPU101 上のプロセスを切り替えるために通常実行される割込みルーチン、プロセススイッチの処理負荷は大きく、CPU101 の負荷が過大となる。

【0086】しかし、本第 1 実施例によれば、割り込み

処理を大幅に削減できるので、結局 CPU にかかる負荷率を大幅に低下できる。

【0087】なお、以上の説明では、I/O 処理装置 2 が、ホスト 1 よりコマンド付加パケットが付加されたコマンドパケットを受け取り、これに基づいて、I/O 処理装置 2 内のメモリ 253 に関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルを作成したが、I/O 処理装置 2 が、ホスト 1 よりコマンドパケットのアドレスにみを受け取り、以降の処理は、このアドレスに基づいて、ホスト 1 内のメモリ 151 内の、コマンドパケット、コマンド付加パケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルに直接アクセスして行なうようにしてもよい。すなわち、メモリ 253 にコマンド付加パケットが付加されたコマンドパケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルを記憶し、これを用いる代わりに、直接、ホスト 1 内のメモリ 151 内の、コマンドパケット、コマンド付加パケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルを用いるようにしてもよい。

【0088】以下、本発明の第 2 の実施例について説明する。

【0089】本第 2 実施例は、第 1 実施例に係る計算機システムにおけるディスク装置 3 がディスクアレイである場合についてのものである。

【0090】ディスクアレイは、複数台のディスクから構成され、これらのディスクにデータを分散し格納するものであるが、ユーザプロセスからは仮想的に 1 台のディスクに見えるようにする必要がある。ユーザプロセスから仮想的に 1 台のディスクに見えるようにするための処理は、ホスト CPU101 が行なう。したがい、ディスクへコマンドを発行する I/O 処理装置 2 はディスクアレイのデータ分散管理には一切関与しない。

【0091】本第 2 実施例に係る計算機システムの構成は、図 1、2、4 に示した前記第 1 実施例に係る計算機システムの構成と同一である。

【0092】ここで、前記第 1 実施例で説明を行なわなかったディスクアレイ管理手段 1042 (図 2 参照) の構成を図 6 に示す。

【0093】図 6 中、10421 は n 台のディスクにデータを分配するためのアドレス変換作業を実施するディスクアレイアドレス変換手段、10422 はディスクアレイの構成情報を格納管理するディスクアレイ情報管理手段、10423 はディスクアレイ中の任意の 1 台のディスク装置が故障してもデータを損失しないように ECC (エラー訂正コード) を生成する ECC 演算手段、10424 はディスクアレイへのコマンド群に関連付けを行うディスクアレイコマンド関連付け管理手段、10425 は ECC 生成を高速化するための ECC およびその生成に必要なデータを保持しておく ECC 生成キャッシュ

ュ手段である。

【0094】以下、本第2実施例に係る計算機システムの動作について説明する。

【0095】ただし、本第2実施例では、前記ECC演算手段10423および前記ECC生成キャッシュ手段10425を用いない場合について説明する。ECC演算手段10423を用いる場合については第3実施例として、ECC生成キャッシュ手段10425を用いる場合については第5実施例として、後に説明する。

【0096】ディスクアクセス要求受信手段1040は、ユーザプロセス102またはディスクキャッシュ管理部106よりファイルアクセス要求を受け取ると、ディスクアレイ全体に対するディスクアクセス要求について親コマンドパケットを生成し、メモリ151に格納し、その格納アドレスをディスクアレイ管理手段1042に渡す。

【0097】ディスクアレイ管理手段1042のディスクアレイアドレス変換手段10421(図6参照)は、この親コマンドパケット受け取り、ディスクアレイ情報管理手段10422の構成情報を参照して、受け取った親コマンドパケットより、個々のディスクへの子コマンドパケット群を生成し、メモリ151に格納する。

【0098】たとえば、ディスクアレイ全体への1つのリードアクセス要求を示す親コマンドパケットが発行されたときに、図7に示すように、そのデータ転送長がiバイトであり、ディスクアレイにおけるデータ分配単位(ストライプサイズ)がkバイトで、 $4k > i > 3k$ の関係にあれば、要求されたデータは4台のディスクに分散することになるので、ディスク1~4の4台への4つの子コマンドパケットを生成する。

【0099】なお、子コマンドパケット群の生成の際には、親、子コマンドパケットに、(1)パケットの階層を示すパケット種、(2)分割コマンド数、(3)子コマンドパケットへのポインタ、(4)自分と同じ階層の次のコマンドパケットへのポインタ、(5)親コマンドパケットへのポインタ等のフィールドを付加し、コマンドパケットを階層的に管理することができるようになる。

【0100】さて、ディスクアレイアドレス変換手段10421は、子コマンドパケット群を生成しメモリ151に格納すると、これらのアドレスをディスクアレイコマンド関連付け管理手段10424に送出する。

【0101】ディスクアレイコマンド関連付け管理手段10424は、これらの子コマンドパケット群のアドレスを受信し、子コマンドパケット群に属する子コマンドパケット間に関連付けを施す。関連付けは、前記第1実施例において説明したディスクコマンド関連付け管理手段10430の行なう関連付けと同じであり、メモリ151上にコマンド付加パケット、関連コマンド管理テーブル1044、コマンドパケットアドレスポインタテーブルを図9に示すとおりに作成することにより行なわれ

る。すなわち、子コマンドパケットは、前記第1実施例における関連コマンドパケットとなる。

【0102】この後、ディスクアレイ管理手段1042はディスクコマンド管理手段1043に各子コマンドパケットのアドレスを渡す。

【0103】ディスクコマンド管理手段1042は、子コマンドパケットのアドレスを受信するが、これらの子コマンドパケット間には、既に関連付けが実施されているので、前記第1実施例で述べたように、ディスク装置管理手段10434での関連付けは行わず、ディスク装置管理手段10434に渡し、ディスク装置管理手段10434に渡されたアドレスのコマンドパケットに、これにリンクしているコマンド付加パケットを付加して、I/O処理装置制御部105よりI/O処理装置に発行する。

【0104】その後の、I/O処理装置2、ディスク装置3の動作は前記第1実施例で説明した動作と同じであり、I/O処理装置2、ディスク装置3は、親コマンドパケットを意識することなく動作する。

【0105】さて、前記第1実施例と同様にしてI/O処理装置2よりの関連コマンドパケット群の終了報告を受けたディスクコマンド管理手段1043は、終了した関連コマンドタグを受信し、ディスクコマンド関連付け管理手段10430で、関連コマンドタグにより、終了した関連コマンドパケットを見つけ、親コマンドパケットへのポインタの存在を判定し、存在していれば、終了報告と関連コマンドタグをディスクアレイ管理手段1042内部のディスクアレイコマンド関連付け管理手段10424に通知する。ディスクアレイコマンド関連付け管理手段10424は、関連コマンドタグにより、終了した関連コマンドパケットを見つけ、親コマンドパケットへのポインタの示す親コマンドパケットに順次リンクしている子コマンドパケット、付加コマンドパケット、関連コマンド管理テーブル、コマンドパケットアドレスポインタテーブルをメモリ151より削除すると共に、親コマンドパケットのフィールド「実行ステータス」に正常終了ステータスを登録後、ディスクアクセス要求終了発行手段1041を介しファイルアクセス要求管理部103に終了報告を行なう。

【0106】このように、本第2実施例によれば、ディスクアレイを用いることで従来の単体ディスクへのアクセスに対し大幅に増加するCPUへの割り込み処理に起因するCPU負荷の上昇を従来の単体ディスク同等に低下させることができる。

【0107】本第3実施例は、前記第2実施例に係る計算機システムにおけるディスクアレイが、ECC機能を有するディスクアレイ、すなわち、RAID (Redundant Arrays of Inexpensive Disks) である場合、すなわち、前記第1実施例におけるディスク装置3がRAIDである場合につい

てのものである。

【0108】なお、RAIDの技術分野ではECCデータのことを一般にパリティと呼ぶので、以下の説明でも、その名称を用いることにする。

【0109】図8に、本第3実施例に係るディスク装置3の構成例を示す。

【0110】図8に示した例は、ディスク数が5台のRAIDについてのものである。5台のディスクのうち、1台はパリティディスクと呼ばれ、ECCデータ専用のディスクである。ディスク0からディスク3までのデータD0, D1, D2, D3の排他的論理和により得られる、パリティデータPが、ディスク4に格納される。すなわち、 $P = D0 + D1 + D2 + D3$ (ただし、本明細書中では、+は排他的論理和を表すものとする) である。

【0111】ここで、ディスク0から4までのディスクの同じディスク内アドレスの領域の集合であるストライプのデータ群をパリティグループと呼ぶ。

【0112】このようなRAIDでは、ディスク1のデータD_iを変更するならば、対応するパリティデータP_iも変更する必要がある。この新しいパリティデータP_iの算出は、新たなデータD_i [new] (新データ) を書き込む以前に、既にディスクに書き込まれているデータD_i [old] (旧データ) とパリティデータP_[old] (旧パリティ) を読みだし、次の演算により新しいパリティデータP [new] (新パリティ) を生成することにより行なわれる。

【0113】 $P [new] = D [new] + D [old] + P [old]$

以上のように、RAIDでは、ディスクアレイのうちの1台のディスクへのライト要求であっても、パリティデータ更新のためのディスクアクセスが発生するので、計4回のディスクアクセスを行なう必要がある。

【0114】したがい、RAIDを用いる場合、CPU101の負荷は、普通の1台のディスクを用いる場合に對し、割り込み処理に要する負荷が4倍に増加することになる。

【0115】以下、本実施例に係る計算機システムの動作について説明する。

【0116】本第3実施例に係る計算機システムの構成は、前記第2実施例に係る計算機システムと同じである。

【0117】さて、本第3実施例では、ディスクアレイ管理手段1042 (図2参照) のディスクアレイアドレス変換手段10421 (図6参照) は、ディスクアレイへのデータD_iのライトを要求する一つのコマンドパケットのアドレスを受信すると、まず、コマンドパケットを参照する。そして、ディスクアレイ情報管理手段10422の構成情報を参照して、ディスクの台数、パリティディスクの位置、ストライプサイズの大きさ等を判断

し、ライトコマンドパケットにより書き込みを要求されたデータを書き込むべきデータの位置はディスク1のD_iであることを決定する。

【0118】次に同様にして、パリティの格納ディスクおよび格納位置を特定する。

【0119】そして、図8のディスク1のデータD_iへのリード要求と、ライト要求と、ディスク4のP_iのリード要求と、ライト要求についての4つのコマンドパケットを生成し、メモリ151に書き込む。なお、この際、旧データおよび旧パリティをロードするメモリ151上の領域、および、新パリティを生成し格納するメモリ151の領域を確保し、そのアドレスを、それぞれのコマンドパケットに登録する。

【0120】次に、ディスクアレイアドレス変換手段10421は、これらのコマンドパケットのアドレスをディスクアレイ関連付け管理手段10424に渡す。

【0121】ディスクアレイ関連付け管理手段10424は、受け取ったコマンドパケット群に対し関連付けを施す。ここで、関連付けは、次のように行なう。

【0122】旧データのリードおよび、旧パリティのリード用のコマンドパケットは、メモリ151上にコマンド付加パケット、関連コマンド管理テーブル1044、コマンドパケットアドレスポインタテーブルを、前記第2実施例と同様に作成することにより関連付ける。これは、両方とも完了しない限り新パリティは生成できないので、両者が終了した時点でI/O処理装置2から終了報告を受ければよいからである。

【0123】また、生成した新パリティのライトと、新データのライトのコマンドパケットも、メモリ151上にコマンド付加パケット、関連コマンド管理テーブル1044、コマンドパケットアドレスポインタテーブルを、前記第2実施例と同様に作成することにより関連付ける。これは両方のライト終了後に、上位にあるファイルアクセス要求手段103へライトアクセス要求の終了報告を行なえばよいので、両者が終了した時点でI/O処理装置2から終了報告を受ければ足りるからである。

【0124】さて、本第3実施例では、次のようにして、ファイルアクセス要求手段103からのディスクアレイアクセス要求に相当するコマンドパケットと、これに基づき作成された2つのリード用のコマンドパケットと2つのライト用のコマンドパケットを関連付ける。

【0125】すなわち、図10に示すように、第2実施例より、階層を1つ増やし、3階層で各コマンドパケット管理するようにする。

【0126】第1の階層 (親コマンドパケット) は、ファイルアクセス要求手段103からのディスクアレイアクセス要求に対応するものである。第2の階層 (子コマンドパケット) は、第1の階層のコマンドパケットがライトについてのものである場合に設けるもので、リード処理に對応する子リードコマンドパケットと、ライト処理に

応する子ライトコマンドパケットが作成される。図示するように、両パケットは親コマンドパケットにリンクしている。

【0127】第3の階層（孫コマンドパケット）は、I/O処理装置2が処理する各々のディスクコマンドをあらわしたものである。2つのリード用孫コマンドパケット（旧データリードコマンドパケット、旧パリティデータリードコマンドパケット）は子リードコマンドパケットにリンクし2つのライト用孫コマンドパケット（新データライトコマンドパケット、新パリティデータライトコマンドパケット）は子ライトコマンドパケットにリンクしている。

【0128】以上のとおり、関連付けが終了したら、ディスクアレイ関連付け管理手段10424は、まず、子リードコマンドパケットにリンクする2つの孫コマンドパケットをディスクコマンド管理手段1043に送出する。

【0129】また、この後、ディスクコマンド管理手段1043は、受信したコマンドパケットのリンクする第1層の親コマンドパケットがライトについてのものである場合には、全てのディスクの、ライトの対象となるストライプに属する領域のロックを設定し、ロックが解除されるまで、受信したコマンドパケットのリンクする第1層の親コマンドパケットにリンクする孫コマンドパケット以外のコマンドパケットの処理を行なわないようする（ディスク資源のロック）。これは、ライト時の旧（パリティ）データのリードと新（パリティ）データのライトは必ず連続して処理を行う必要があるためである。

【0130】ディスクコマンド管理手段1043は、この後、これらの孫コマンドパケットにリンクする付加コマンドパケットを付加してディスクコマンド発行手段1050を渡し、ディスクコマンド発行手段1050該部1050は、これをI/O処理装置2に発行する。

【0131】I/O処理装置2では、前記第1実施例と同様に、2つのリード用コマンドパケットを処理し、両者が終了すると、ホスト1のOS1010に一度の終了割り込を発行する。前記第1実施例と同様に、これにより起動されるディスクコマンド終了受信手段1051はI/O処理装置2から終了した関連コマンドの関連コマンドタグを獲得し、ディスクコマンド管理手段1043に渡す。ディスクコマンド管理手段1043は、前記第2実施例と同様に、これをディスクアレイ管理手段1042に渡す。

【0132】さて、ディスクアレイ管理手段1042のディスクアレイ関連付け管理手段10424は、受け取った関連コマンドタグより、処理が終了した子パケットを判定する。そして、子リードパケットの処理が終了したことを確認したならば、子リードパケットの実行ステータスに正常終了を登録し、ECC演算手段10423

に処理を渡す。

【0133】ECC演算手段10423はこれを受け、処理が終了した子リードパケットに親コマンドパケット、子ライトコマンドパケットを介してリンクしているライト用孫コマンドパケットのうちの書き込みデータについてのものから新データの格納されているメモリアドレスと、処理が終了した子リードパケットにリンクしている2つのリード用孫コマンドパケットから旧データ、旧パリティデータの格納されているメモリアドレスを獲得する。

【0134】そして、これら3データをメモリ151より読み出し、排他的論理和を演算し、メモリ151の、新パリティライト用の孫コマンドパケットに記述されている格納アドレスに、排他的論理和演算結果を格納する。

【0135】ECC演算手段10423は、この処理のあと、処理を再びディスクアレイ管理手段1042に戻す。ディスクアレイ管理手段1042は新パリティの生成が終了したので、子ライトコマンドパケットにリンクする2つのライト用孫コマンドパケットのアドレスをディスクコマンド管理手段1043に送出する。

【0136】ディスクコマンド管理手段1043は、ディスクコマンド発行手段1050を介して、I/O処理装置2に2つのライト用孫コマンドパケットを発行する。

【0137】I/O処理装置2では、前記第1実施例と同様に、2つのライト用コマンドパケットを処理し、両者が終了すると、ホスト1のOS1010に一度の終了割り込を発行する。前記第1実施例と同様に、これにより起動されるディスクコマンド終了受信手段1051はI/O処理装置2から終了した関連コマンドの関連コマンドタグを獲得し、ディスクアレイ管理手段1042に渡す。

【0138】ディスクコマンド管理手段1043は、前記第2実施例と同様に、これをディスクアレイ管理手段1042に渡す。また、ディスクコマンド管理手段1043は、受け取った関連コマンドタグに対応するライト用孫コマンドパケットが、ライトを要求する第1層の親コマンドパケットにリンクするものである場合には、この親コマンドについて行なったロックの設定を解除する。

【0139】一方、ディスクアレイ管理手段1042のディスクアレイ関連付け管理手段10424は、受け取った関連コマンドタグより、処理が終了した子パケットを判定する。そして、子リードパケットの処理が終了したことを確認したならば、子ライトコマンドパケットの実行ステータスに正常終了を登録し、子ライトコマンドパケットにリンクする親コマンドパケットにリンクする子リードコマンドパケットの実行ステータスが正常終了であることを確認し、親コマンドパケットの実行ステ

タスに正常終了ステータスを登録後、ディスクアクセス要求終了発行手段1041を介しファイルアクセス要求管理部103に終了報告をし、すべての子コマンドパケット、孫コマンドパケット、コマンド付加パケット、関連コマンド管理テーブル、コマンドパケットアドレスポインターテーブルを削除する。

【0140】一方、ファイルアクセス要求管理部103は、親コマンドパケットの正常終了ステータスを確認後、ファイルアクセス要求元に正常終了報告を行ない、親コマンドパケットをメモリ151より削除する。

【0141】以上のように、本第3実施例によればRAIDをディスク装置として用いる場合に、ライト時に行なわれる複数のディスクアクセスに関連付けを施し、CPU割り込み処理の増加を防ぐことができる。

【0142】以下、本発明の第4の実施例について説明する。

【0143】本第4実施例は、前記第1実施例においてディスクキャッシュ部106を使用した場合についてのものである。

【0144】本第4実施例に係る計算機システムの構成を前記第1実施例に係る計算機システムの構成と同じである。

【0145】本第4実施例では、ファイル要求管理部106は、ディスクからデータをリードした際にディスクキャッシュ部106に、そのデータを保管し、もし、つぎに同じデータのリード要求がユーザプロセス102より発行された際には、ディスクアクセスを行なうことなく主記憶のキャッシュ用領域からデータをリードし要求元のユーザプロセスに与える。また、ライト時にディスクキャッシュ部106にデータを書き込み、ライト要求元のユーザプロセス102に終了通知を発行する。また、ファイル要求管理部106には、使用状況を管理する管理テーブルを設ける。

【0146】さて、このようにして、ユーザプロセスよりのライト要求に応じてディスクキャッシュ部106に書き込んだデータは、ある時間後、まとめてディスク装置3にライトする必要がある。

【0147】以下、このディスクキャッシュ部106のデータのディスク装置3への書き出し動作について説明する。

【0148】いま、図11に示すように、メモリ151上のディスクキャッシュ部106には、A、B、C、D、Eのデータブロックがディスクに書き込まれていない状態（これをダーティブロックと呼ぶ）で保持されている。このダーティブロックは適当な時間間隔でディスクに書き出される。この書き出しのタイミングの決定の方式としては、いくつかの方式が知られているが、本第4実施例では、一定時間毎に検査し、もしダーティブロックが存在するならばこれらを全て書き出すものとして説明する。

【0149】さて、ディレーライトキャッシュデータライト要求部1020は一定時間毎にダーティキャッシュ書きだし命令をファイルアクセス要求管理部103に送出する。ファイルアクセス要求管理部103はこれを受信し、ディスクキャッシュ部106の使用状況を管理する管理テーブルを操作し、ダーティブロックを検出する。

【0150】ファイルアクセス要求管理部103は、これら検出したダーティブロックのディスクライトアクセス要求を作成し、順次、ディスクコマンド管理部104に渡す。

【0151】ディスクアクセス要求手段1040は、これらのディスクアクセス要求を受信し、対応するコマンドパケットを生成してメモリ151上に書き込み、そのアドレスを、ディスクコマンド管理手段1043に送出する。ディスクコマンド管理手段1043では、渡されたアドレスのコマンドパケットを読み出し、前記第1実施例と同様に、ディスクコマンド関連付け機能サポート有無判定手段10431でI/O処理装置2の機能を確認し、ディスクコマンド関連既所持有無判定手段10432で関連既所持のないことを判定し、キューイング手段10433に、そのアドレスを格納する。このようなダーティブロックの書き出しは一度に多数のコマンドパケットを同時に発生させるので、キューイング手段10433には多数のコマンドパケットのアドレスが格納される。

【0152】ディスクコマンド関連付け管理手段10430はキューイング手段10433を検査し、アドレスが存在する全てのライトコマンドパケットに、またはある規定個数のライトコマンドパケットに、または、複数のディスクがI/O処理装置2に接続している場合には各ディスクにつき1つずつ抽出したコマンドパケットに関連付けを施す。

【0153】ここで、ライトキャッシュを行なっているディスクキャッシュを持つホストでは、ユーザのライト要求はディスクキャッシュ部106に書き込んだ時点で終了しており、関連付けにより一つ一つのダーティブロックライトの書きだし処理終了が若干遅延しても問題はほとんど発生しないので、このように多数のライトコマンドパケットに関連付けを施しても問題は生じない。しかし、まれにキャッシュあふれ等により即時に処理を実施したいという要求がある場合には、ファイルアクセス要求管理部103からの指定に応じて、コマンドパケットに関連付けを釣り支するフィールドを設け、このようなコマンドパケットについては関連付けの対象としないようにしてもよい。

【0154】以下の動作は、前記第1実施例と同じである。

【0155】なお、本実施例は、前記第2、第3実施例と組み合わせて適用することもできる。すなわち、第

2、第3実施例のディスクアレイ管理手段1042において、複数の親コマンドパケットにリンクする、複数のリード用孫コマンドパケット間、および、複数のライト用孫コマンドパケット間で関連付けを施すようにしてもよい。

【0156】以上のように、本第4実施例によれば、ライトキャッシュを有するホスト1において、CPUの負荷を低減することができる。

【0157】以下、本発明の第5実施例について説明する。

【0158】本第5実施例は、前記第3実施例において、ディスクアレイ管理手段1042のECC(パリティ)生成キャッシュ手段10425を用いる場合についてのものである。

【0159】前記第3実施例では新パリティを生成する際に、すべての必要なデータ(旧データ、旧パリティ)をディスクから読みだした後に、新パリティを計算し、パリティディスクに新データと共に送出していた。しかし、これでは、1回のライトにつき必ず2台のディスクを占有することになる。

【0160】そこで、本第5実施例では、次のようにして、一度に占有するディスクを1台とすることにより、並列度を高める。

【0161】すなわち、まず、旧データをリードして新データをライトし、新旧データの排他的論理和を計算し、両データの差分データまでを生成し、メモリ151に保存する。そして、適当な遅延時間後に旧パリティをリードして旧パリティと上記差分データとの排他的論理和を計算し新パリティ生成してライトする。

【0162】また、本第5実施例では、より高速化するために、旧データを毎回読みだすのではなく、前記ECC生成キャッシュ手段10425が管理するメモリ151上の所定の領域に、リードしたデータを保持しておく。このようにすることにより、旧データとして、このデータを用いることができる可能性がある。この場合、旧データを読みだす必要がなくなる。また、さらに、旧パリティも毎回読み出すのではなく、前記ECC生成キャッシュ手段10425が管理するメモリ151上の所定の領域に、リードしたパリティを保持しておく。これにより、旧パリティとして、このデータを用いることができる可能性がある。また、ディスクに書き込む新データや新パリティを保持しておく。これにより、その後、旧データ、旧パリティとして、これらを用いることができる可能性がある。

【0163】なお、このようなデータやパリティの保持は前記第4実施例で示したディスクキャッシュ部106で行なうようにしてもよい。

【0164】さて、ファイルアクセス要求管理部103より、ライトアクセス要求のコマンドパケットを受け取ると、ディスクアドレス変換手段10421は、図12

に示すように、これを親コマンドパケット#1とし、これをコピーすることによりメモリ151上に作成した親コマンドパケットを#2とする。また、図示するように、親コマンドパケットに、NEXTという名称のフィールドを設け、これを用いて、親コマンドパケット#1と親コマンドパケット#2をリンクする。

【0165】そして、次に、親コマンドパケット#1の子コマンドパケットとして、旧データをリードするための子コマンドパケット#1と、旧データをライトするためのコマンドパケット#2を生成し、親コマンドパケット#2の子コマンドパケットとして、旧パリティをリードするための子コマンドパケット#3と、旧パリティをライトするためのコマンドパケット#4を生成し、4つの子コマンドパケットを生成してメモリ151に格納し、これらのアドレスを、ディスクアレイ関連付け管理手段10424に渡す。

【0166】ディスクアレイ関連付け管理手段10424は、まず、旧データをリードするためのコマンドパケット#1を参照し、この旧データをECC生成キャッシュ手段10425管理しているメモリ151上の領域に存在しているかを判定し、保持していれば、この子コマンドパケット#1の実行ステータスを正常終了とし、終了報告をディスクアドレス変換手段10421に行なう。ECC演算手段10423では、旧データをリードするためのコマンドパケット#1に親パケット#1、#2を介してリンクしている新パリティライト用のコマンドパケット#4を参照し、フィールド「メモリアドレス」にアドレスが登録されていない場合には、旧データと、新データライト用のコマンドパケットが指定する新データとの排他的論理和(差分データ)を計算して、ECCキャッシュ生成手段10425を介してメモリ151に格納し、そのメモリアドレスを新パリティライト用のコマンドパケット#4のフィールド「メモリアドレス」に登録する。

【0167】もし、旧データを保持していないければ、この子コマンドパケット#1はディスクコマンド管理手段1043に渡され、ディスクに送出し、旧データを読みだし、処理が終了した旧データリード用のコマンドパケットの実行ステータスを正常終了とし、ディスクアドレス変換手段10421に終了報告を行なう。ECC演算手段10423は、先程の場合と同様に、旧データをリードするためのコマンドパケット#1に親パケット#1、#2を介してリンクしている新パリティライト用のコマンドパケット#4を参照し、フィールド「メモリアドレス」にアドレスが登録されていない場合には、旧データと、新データライト用のコマンドパケットが指定する新データの排他的論理和を計算して、ECC生成キャッシュ手段10425を介してメモリ151に格納し、そのメモリアドレスを新パリティライト用のコマンドパケット#4のフィールド「メモリアドレス」に登録す

る。

【0168】次に、ディスクアレイ関連付け管理手段10424は、新データライト用のコマンドパケット#2をディスクコマンド管理手段1043に渡す。ディスクコマンド管理手段1043は、前記第4実施例と同様、他のライトコマンドパケットとの間で関連付けを施された後、ディスクに送出される。以降、前記第1実施例と同様に処理が行なわれ、すべての関連コマンドパケットに対応するデータ転送が終了したならば1回の割り込みがI/O処理装置から発行され、ディスクアレイコマンド関連付け管理手段10424に処理が戻り、ディスクアレイコマンド関連付け管理手段10424は、処理が終了した新データライト用のコマンドパケットの実行ステータスを正常終了とし、ディスクアドレス変換手段10421に終了報告を行なう。ディスクアドレス変換手段10421は、これより親コマンドパケット#1の終了を認識し、親コマンドパケット#1の実行ステータスを正常終了とし、ファイルアクセス管理部103に終了報告を行う。ファイルアクセス管理部103は、ライトアクセスの要求元のユーザプロセス102に終了報告を行なう。

【0169】なお、実際には、この時点では、パリティの更新が終了していないので、ユーザプロセス102への終了報告は、パリティの更新に先行して行なわれることになる。

【0170】次に、ディスクアレイ関連付け管理手段10424は、旧パリティリードのための子コマンドパケット#3を参照して、旧パリティがECC生成キャッシュ手段10425が管理するメモリ151上の領域に存在しているかを判定し、存在すれば、このコマンドパケットの実行ステータスを正常終了とし、終了報告をディスクアドレス変換手段10421に行なう。また、ECC演算手段10423は、旧パリティリードのための子コマンドパケット#3に親コマンドパケット#2を介してリンクしている新パリティのライトのための子コマンドパケット#4のフィールド「メモリアドレス」に登録されたアドレスに登録されている旧データと新データの差分データをECC生成キャッシュ手段10425を介して読み出し、これと、獲得した旧パリティの排他的論理和を計算し、子コマンドパケット#4のフィールド「メモリアドレス」に登録されたメモリ151のアドレスに書き込む。そして、ECC生成キャッシュ手段10425内部のキューイング手段に、この子コマンドパケット#4をキューイングする。

【0171】また、もし、旧パリティをECC生成キャッシュ手段10425が保持していない場合は旧パリティ用、新パリティ用の両方の子コマンドパケット#3、#4を前記ECC生成キャッシュ手段10425のキューイング手段にキューイングする。

【0172】さて、適当な時間経過後、ECC生成キャ

ッシュ手段10425のキューイング手段内のコマンドパケットの発行を行なう。すなわち、ディスクアレイコマンド関連付け管理手段10424は、ECC生成キャッシュ手段10425のキューイング手段を走査し、適当な個数の旧パリティリード用子コマンドパケット#3のみに関連付けを施し、子コマンドパケット#3をディスクコマンド管理手段1043に送出する。以降、前記第1実施例と同様に処理が行なわれ、すべての関連コマンドパケットに対応するデータ転送が終了したならば上記第1、2、3、4の実施例同様1回の割り込みがI/O処理装置から発行され、ディスクアレイコマンド関連付け管理手段10424に処理が戻り、ディスクアレイコマンド関連付け管理手段10424は、処理が終了した旧パリティリード用のコマンドパケットの実行ステータスを正常終了とし、ディスクアドレス変換手段10421に終了報告を行なう。

【0173】さて、このようにして、旧パリティ用のリードコマンドパケット#3が正常終了して旧パリティが得られたら、ECC演算手段10423は、これを用いて新パリティを得て、ディスクアレイコマンド関連付け管理手段10424は、キュー中の、新パリティが得られた新パリティ用の子コマンドパケット#4を複数個の関連付ける。そして、関連付けた新パリティ用の子コマンドパケットをディスクコマンド管理手段1043に送出する。

【0174】以降、前記第1実施例と同様に処理が行なわれ、すべての関連コマンドパケットに対応するデータ転送が終了したならば上記第1、2、3、4の実施例同様1回の割り込みがI/O処理装置から発行され、ディスクアレイコマンド関連付け管理手段10424に処理が戻り、ディスクアレイコマンド関連付け管理手段10424は、処理が終了した新パリティライト用のコマンドパケットの実行ステータスを正常終了とし、ディスクアドレス変換手段10421に終了報告を行なう。

【0175】さて、このようにして、一つの親コマンドパケットに対応する4つの子コマンドパケットの終了報告が揃ったら、ディスクアドレス変換手段10421は、前記第2実施例と同様に、その親コマンドパケット#2の実行ステータスを正常終了とし、終了処理を行なう。ここで、ファイルアクセス管理部103、ライトアクセスの要求元のユーザプロセス102への終了報告は、既にすんでいるので行なわない。

【0176】以上のように、本第5実施例によれば、パリティの更新をデータの更新と分離し、パリティを遅延させて更新するような場合に、ECC生成のキャッシュ利用による高速化により、ディスクアクセス性能を高速化することができると共に、CPUの割り込み処理を大幅に削減することができる。

【0177】なお、以上の各実施例では、I/O処理装置2において、関連コマンド終了判定手段2031はす

べての関連付けられたコマンドが終了するまでホスト1に終了報告を行なわないが、このようにすると、もし、関連付けられたコマンドの一つにエラーが生じた場合に、永遠に終了報告が延期されてしまう可能性がある。そこで、関連コマンド終了判定手段2031はあらかじめ設定した適当な時間でタイムアウトとし、関連付けられたコマンドコマンドの終了・未終了にかかわらずホストに終了報告を発行するようにしてよい。

【0178】ホスト1のディスクコマンド管理部104はこの終了報告を受信すると、関連コマンドのステータスを確認し、未終了のコマンドがあったことを確認し、再度コマンドを発行するか、もしくは、コマンド実行失敗として要求元に報告する。

【0179】また、以上の各実施例は、ディスク装置ではなく、光ディスク装置、半導体記憶装置、磁気テープ装置、通信制御装置、等他のI/O処理を行う場合にも適用することができる。

【0180】

【発明の効果】以上のように、本発明によれば、ディスクアクセスに起因する、CPUへの割り込み発生回数を削減することにより、CPUの負荷を低減することのできる計算機システムを提供することができる。

【図面の簡単な説明】

【図1】本発明の第1実施例に係る計算機システムの構成を示すブロック図である。

【図2】本発明の実施例に係る計算機システムのディスクアクセスコマンド処理に関する部位の構成を示したブロック図である。

【図3】本発明の第1実施例において用いるコマンドパケット、関連コマンド管理テーブルの構造および両者の関係を示した図である。

【図4】本発明の実施例に係るディスクコマンド管理手段の構成を示すブロック図である。

【図5】本発明の実施例に係る計算機システムのディスクアクセスの動作を示すタイムチャートである。

【図6】本発明の実施例に係るディスクアレイ管理手段の構成を示すブロック図である。

【図7】ディスクアレイの複数のディスクと、アクセス

要求の関係を示す図である。

【図8】RAIDにおけるデータとパリティの関係を示す図である。

【図9】本発明の第2実施例において用いるコマンドパケットと関連コマンド管理テーブルの構造および両者の関係を示した図である。

【図10】本発明の第3実施例において用いるコマンドパケットと関連コマンド管理テーブルの構造および両者の関係を示した図である。

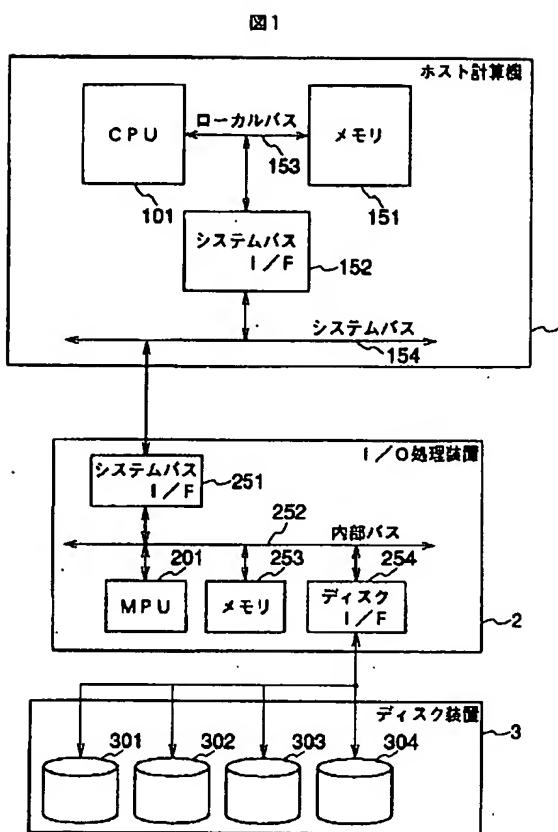
【図11】本発明の第4実施例におけるディスクアクセスコマンドの経路を示す図である。

【図12】本発明の第5実施例において用いるコマンドパケットと関連コマンド管理テーブルの構造および両者の関係を示した図である。

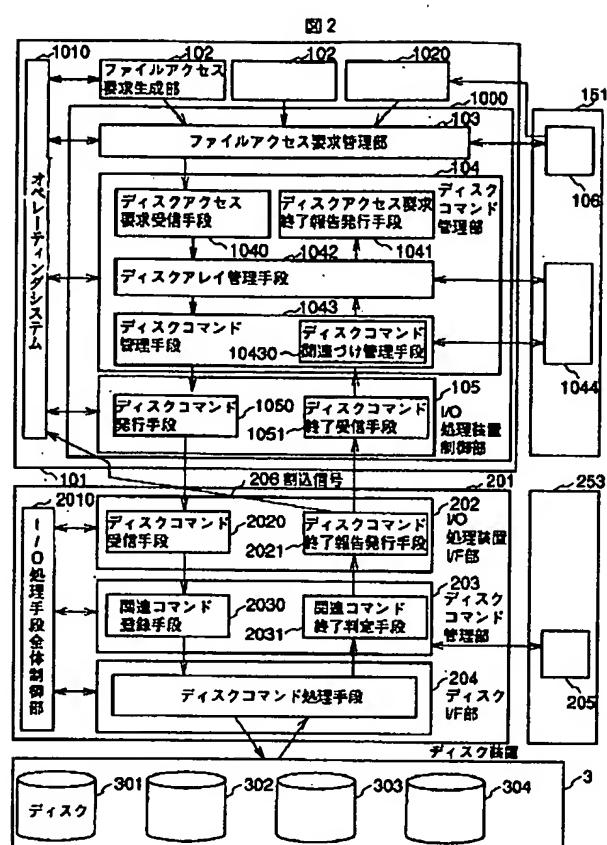
【符号の説明】

- 1・・・ホスト計算機
- 2・・・I/O処理装置
- 3・・・ディスクアレイ装置（又はディスク装置）
- 101・・・CPU
- 102・・・ファイルアクセス要求生成部
- 103・・・ファイルアクセス要求管理部
- 104・・・ディスクコマンド管理部
- 105・・・I/O処理装置制御部
- 151・・・メモリ
- 201・・・MPU
- 202・・・I/O処理装置I/F部
- 203・・・ディスクコマンド管理部
- 204・・・ディスクコマンド処理部
- 205・・・コマンド管理エリア
- 253・・・メモリ
- 1000・・・ドライバ部
- 1010・・・オペレーティングシステム
- 1042・・・ディスクアレイ管理手段
- 1043・・・ディスクコマンド管理手段
- 1044・・・コマンド管理エリア
- 2030・・・関連コマンド登録手段
- 2031・・・関連コマンド終了判定手段
- 10430・・・ディスクコマンド関連付け手段

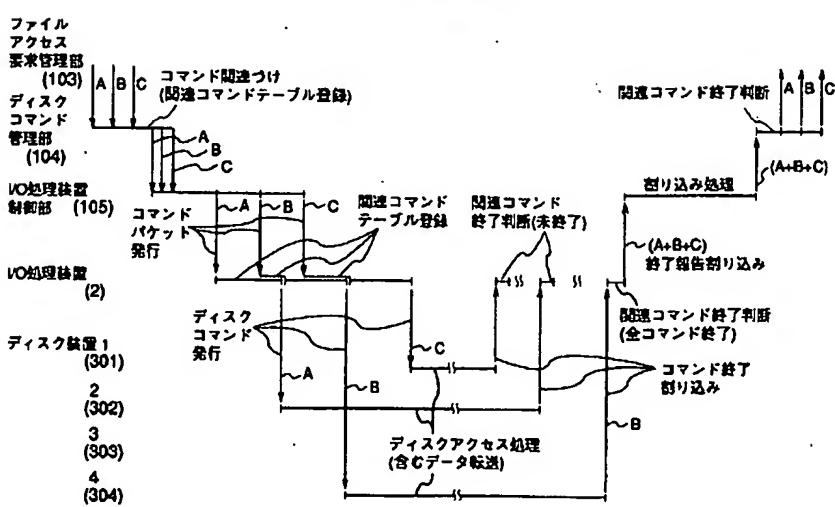
【図1】



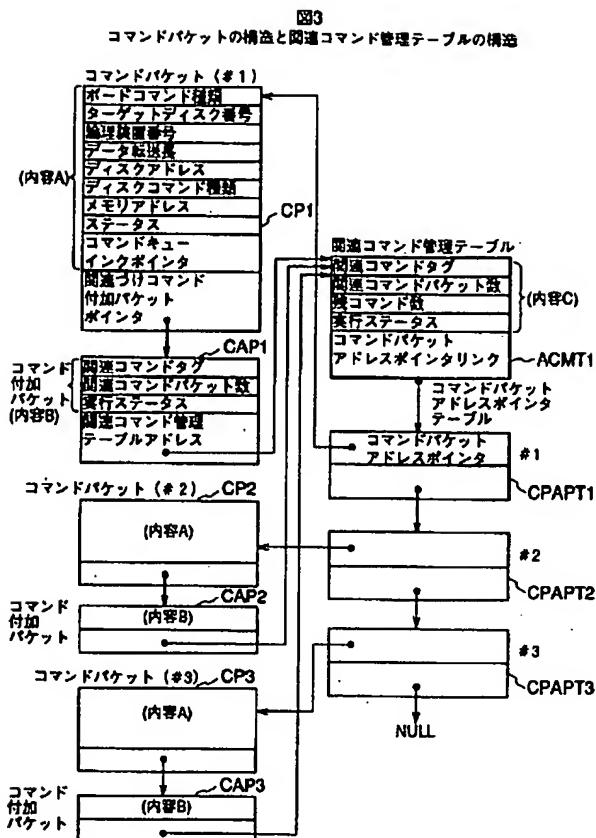
【図2】



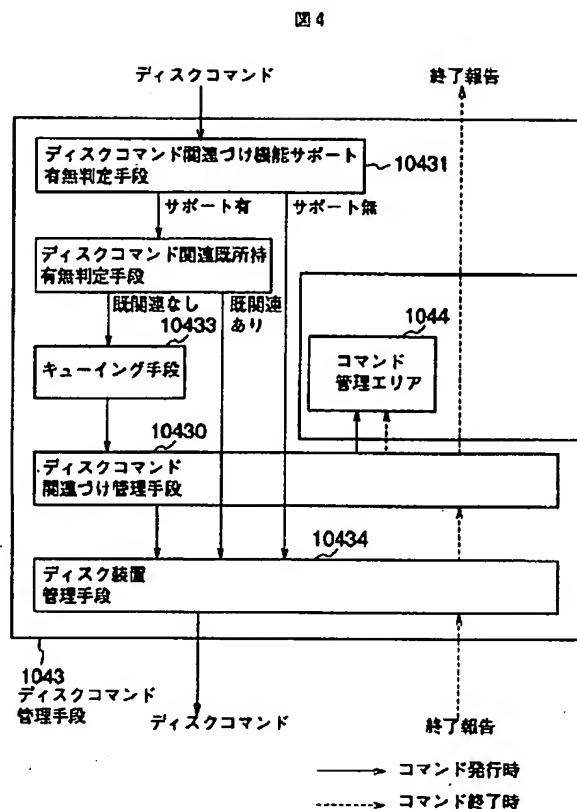
【図5】

図5
コマンド管理動作図

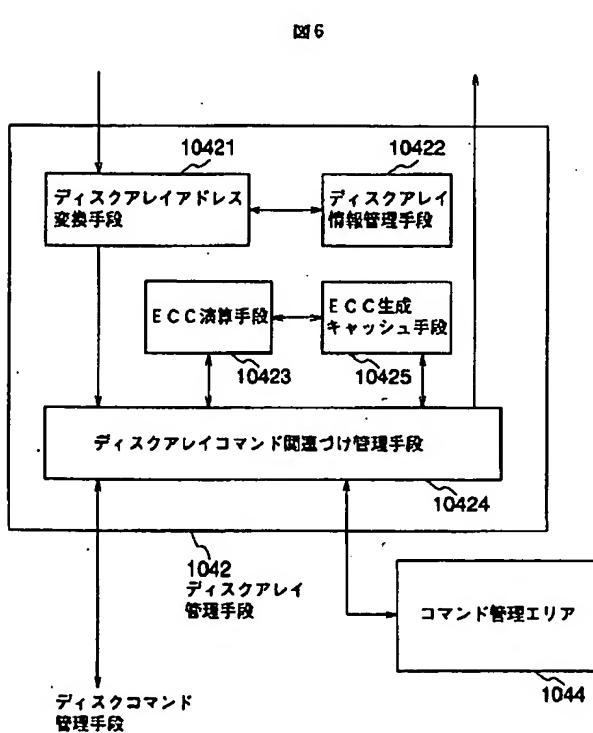
【図3】



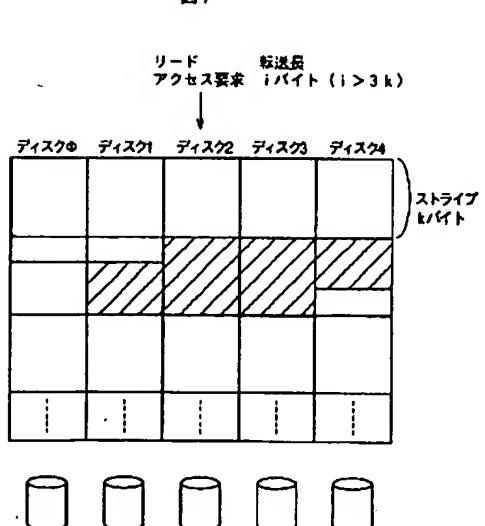
【図4】



【図6】



【図7】



[图 8]

[图 9]

8

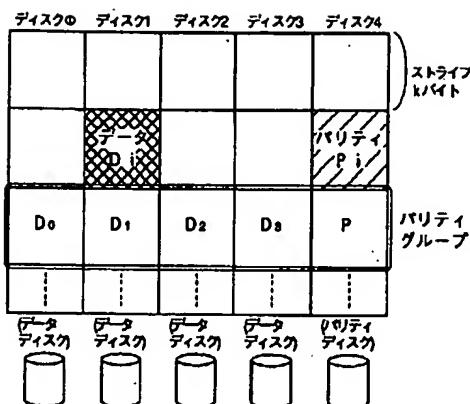
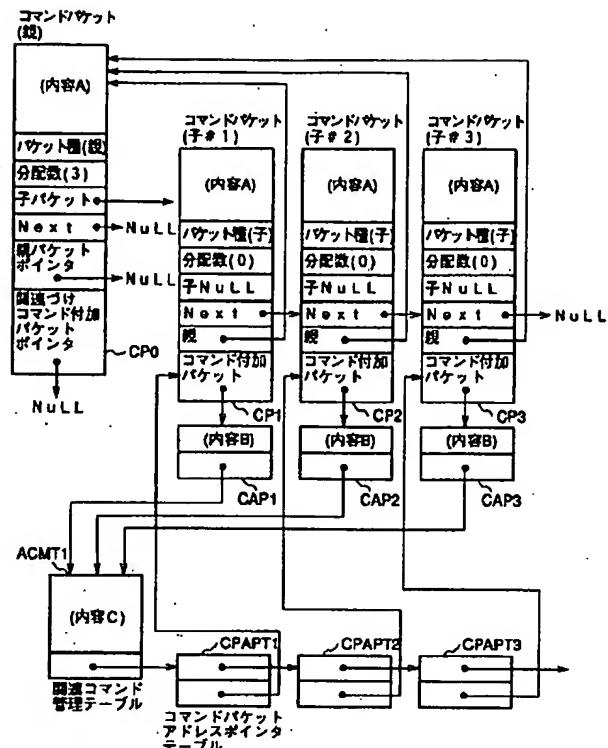
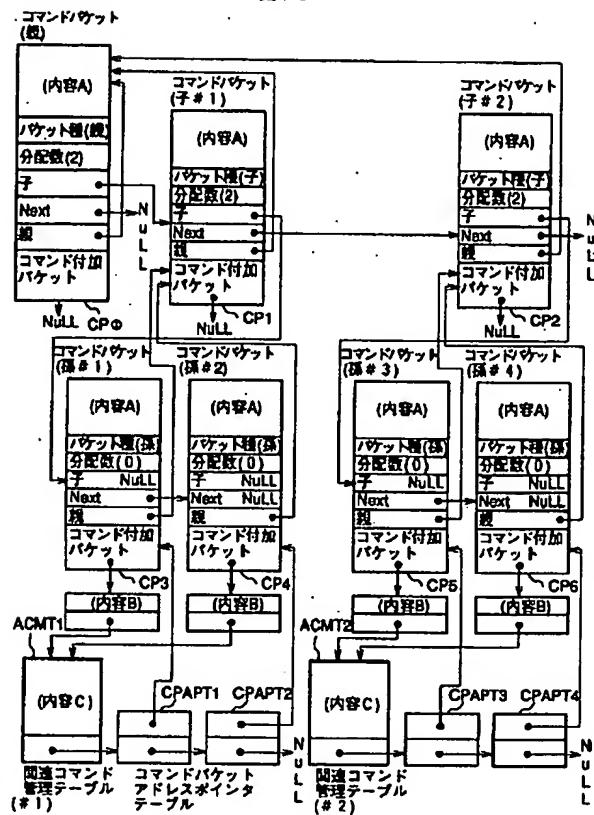


图 9



【図10】

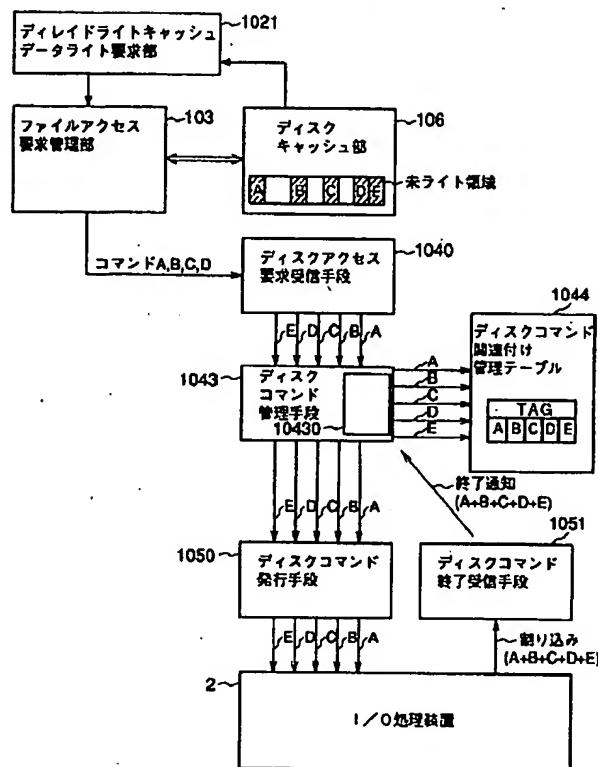
図10



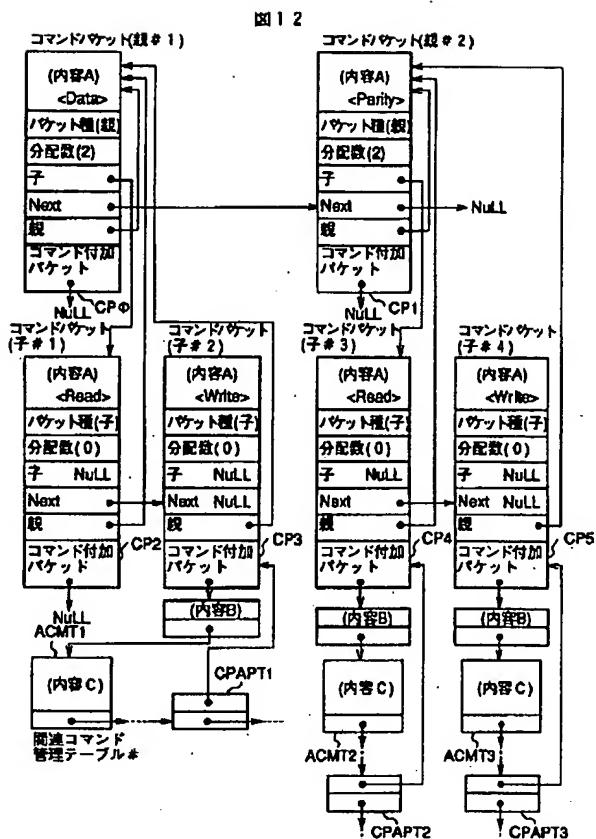
【図11】

図11

ライトキャッシュと間連づけ



【図12】



フロントページの続き

(72) 発明者 八木沢 育哉

神奈川県川崎市麻生区王禅寺1099 株式会
社日立製作所システム開発研究所内

(72) 発明者 大枝 高

神奈川県川崎市麻生区王禅寺1099 株式会
社日立製作所システム開発研究所内

(72) 発明者 荒川 敬史

神奈川県川崎市麻生区王禅寺1099 株式会
社日立製作所システム開発研究所内

【公報種別】特許法第17条の2の規定による補正の掲載
 【部門区分】第6部門第3区分
 【発行日】平成11年(1999)7月30日

【公開番号】特開平7-271521
 【公開日】平成7年(1995)10月20日
 【年通号数】公開特許公報7-2716

【出願番号】特願平6-57197

【国際特許分類第6版】

G06F	3/06	540
	13/10	340

【F I】

G06F	3/06	540
	13/10	340 B

【手続補正書】

【提出日】平成10年7月1日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】請求項9

【補正方法】変更

【補正内容】

【請求項9】請求項8記載の計算機システムであって、前記ホスト装置のCPUは、前記ディスクアレイより読み出したデータおよびパリティを、メモリ上に設けたキャッシュ領域に保持する手段を有し、前記関連付け手段は、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきデータに対応するデータが存在する場合には、現データ読み出し用コマンドを前記入出力処理装置に渡さず、前記キャッシュ領域に、現パリティ読み出し用コマンドによって読み出すべきパリティに対応するパリティが存在する場合には、現パリティ読み出し用コマンドを前記入出力処理装置に渡さず、前記パリティ生成部は、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきデータに対応するデータが存在する場合には、前記ディスクアレイより読み出した前記現データに代えて前記キャッシュ領域に存在するデータを用いて、前記現データと新データとの排他的論理和を生成し、前記キャッシュ領域に、現データ読み出し用コマンドによって読み出すべきパリテ

ィに対応するパリティが存在する場合には、前記ディスクアレイより読み出した前記現パリティに代えて前記キャッシュ領域に存在するパリティを用いて、現パリティと保持したパリティの排他的論理和を、前記新パリティ書き込み用コマンドによって書き込む新パリティとして生成することを特徴とする計算機システム。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】請求項10

【補正方法】変更

【補正内容】

【請求項10】ホスト装置と補助記憶装置との間の入出力を担う入出力処理装置であって、ホスト装置より渡された複数のコマンドに付加されている、当該コマンドの属するグループを示す情報に基づいて、渡された各コマンドと当該コマンドが属するグループの対応を表す情報を作成して記憶する手段と、ホスト装置より渡された各コマンドの指示する補助記憶装置のアクセスを実行する手段と、記憶した各コマンドと当該コマンドが属するグループの対応に基づいて、1つのグループに属する全てのコマンドの指示するアクセスが終了した場合に、当該グループについての処理が終了したことを前記ホスト装置に割り込みを用いて報告する手段とを有することを特徴とする入出力装置。

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.